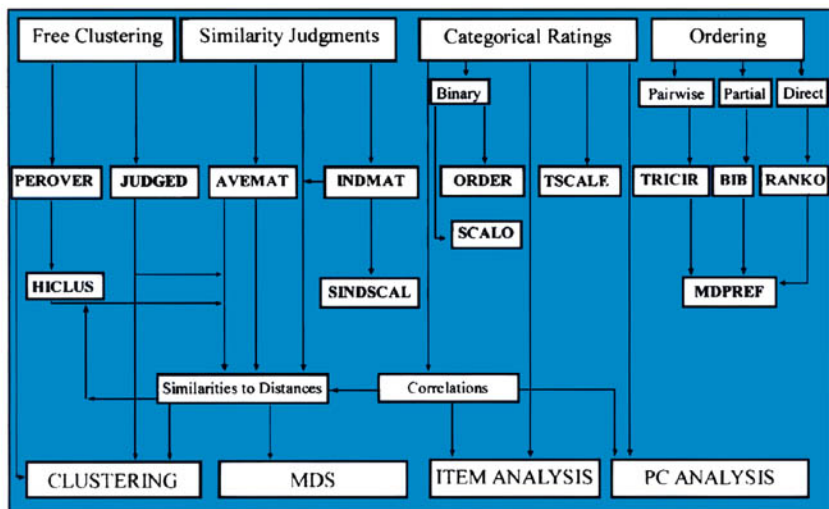


SCALING METHODS

SECOND EDITION



Peter Dunn-Rankin
Gerald A. Knezek
Susan Wallace
Shuqiang Zhang

SCALING METHODS

2nd Edition

This page intentionally left blank

SCALING METHODS

2nd Edition

Peter Dunn-Rankin

University of Hawaii at Manoa

Gerald A. Knezek

University of North Texas

Susan Wallace

University of North Florida

and

Shuqiang Zhang

University of Hawaii at Manoa



2004

LAWRENCE ERLBAUM ASSOCIATES, PUBLISHERS
Mahwah, New Jersey London

Copyright © 2004 by Lawrence Erlbaum Associates, Inc.

All rights reserved. No part of this book may be reproduced in any form, by photostat, microform, retrieval system, or any other means, without prior written permission of the publisher.

Lawrence Erlbaum Associates, Inc., Publishers
10 Industrial Avenue
Mahwah, NJ, 07430

Cover design by Kathryn Houghtaling Lacey

Library of Congress Cataloging-in-Publication Data

Scaling methods.-- 2nd ed. / Peter Dunn-Rankin ... [et al.].
p. cm.

Rev. ed. of: Scaling methods / Peter Dunn-Rankin.

Includes bibliographical references and index.

ISBN 0-8058-1802-2

1. Scale analysis (Psychology) I. Dunn-Rankin, Peter. II. Dunn-Rankin, Peter.

Scaling methods.

BF39.2.S34S33 2004

150'.28'7--dc22

2003049460

Books published by Lawrence Erlbaum Associates are printed on acid-free paper, and their bindings are chosen for strength and durability.

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

Disclaimer:

This eBook does not include the ancillary media that was packaged with the original printed version of the book.

CONTENTS

PREFACE

xi

What's New? xi
Content and Organization xiii
Acknowledgements xiii

PART I: FOUNDATIONS

1

1. SCALING DEFINED

3

Relative Measurement 3

The Fahrenheit Scale 3
Psychological Objects 3
Mapping 4
Introduction to Scaling 4
Euclidean Space 8
Guttman Scales 9
Judgments or Choices 9

2. TASKS

11

Ordering 11

Paired Comparisons 12
Circular Triads 12
Partial Ranks and Balanced Incomplete Block Designs 12
Direct Ranking 14
Ranks and Rank Values 14
Tetrads (Pairs of Pairs) 14
Arranging Pairs 15
Flow Diagram for Analysis of Ordinal Tasks 15
CD-ROM Example of BIB 17

Categorical Ratings 18

Judgments 18
The Semantic Differential 19
Simple Scoring 19
Subsets of Items 20
Steps in Ordered Category Scale Construction 20
Ordered Category Example 20

Restrictions of Ordered Categories	21
Number of and Naming of Categories	21
Flow Diagram for Ordered Category Analysis	22

Free Clustering 23

Steps in Free Clustering	23
Inter-Judge Distances	24
Individualized Free Clustering	25
Flow Diagram for Free Clustering Analysis	25
CD-ROM Example of Using PEROVER	26
CD-ROM Example of Using JUDGED	27

Similarity Judgments 27

Paired Comparisons	27
Ranking Pairs	29
Rating Similarity Between Pairs	29
Clustering Then Pairing	30
Triadic Comparisons	30
Ratio Estimation	32
Conditional Ranking	32
Same-Different	33
Latency	33
Ranking Versus Rating Pairs	33
Analysis of Similarities	34
Flow Diagram for Similarity Judgments Analysis	35
CD-ROM Example of AVEMAT	35
CD-ROM Example of INDMAT	36

3. MEASURES OF PROXIMITY

37

Correlations 37

Pearson's Correlation	37
SAS Example of Calculating Correlations	38
Significance of r	39
Squaring the Correlation Coefficient	40
Kendall's tau Correlation	40
Gamma Correlation	42

Distances 42

Standardized Distance	42
Mahalanobis d^2	43
Minkowski Metric	43
Triangle Inequality	44

Scalar Products 45

Association 47

Direct Estimation of Proximity	47
Percent Overlap	47
Minimum Percentage	48
Interjudge Distances Following Free Clustering	49

Gower's Similarity Measure	49
Kappa	51
A Distance Macro from SAS	52

PART II: UNIDIMENSIONAL METHODS **53**

4. RANK SCALING **55**

Variance Stable Rank Sums **55**

Test of Significance	57
Number of Judges	58
Discussion	59
Application 1: Direct Ranking of Counselor Roles	60
Application 2: Letter Similarity Scales	62
CD-ROM Example Using RANKO	64

Circular Triad Analysis **66**

Judge Circular Triads (JCT)	66
Coefficient of Consistency	67
Tests for Circularity	67
Application: Circularity Among Adjective Pairs	68
Circular Triad Analysis	69
Discussion	71
CD-ROM Example Using TRICIR	71

5. ORDER ANALYSIS **75**

Guttman Scaling **75**

Goodenough's Error Counting	76
Application 1. Cloze Tests in Reading	79
Application 2. Arithmetic Achievement	80
Significance of a Guttman Scale	80
CD-ROM Example Using SCALO	81

Mokken Scales **80**

Dominance Theory of Order **83**

CD-ROM Example Using ORDER	87
Fisher's Exact Probability	88
CT3 Index	89
Rescaling Reliability	90
Application Example	90

Partial Correlations As A Measure of Transitivity **91**

6. COMPARATIVE JUDGMENT **93**

Attitudes are Normally Distributed	93
Thurstone's Case V	94
Case V Example	95
Reliability	97

Application: Seriousness of Crimes Then and Now	97
Case V Program	97

7. CATEGORICAL RATINGS **99**

Greens' Successive Categories 100

Discussion	103
TSCALE Analysis of Reading Attitude	104

Summated Ratings 105

An Example of Likert Scaling	105
Discussion	106
Example: Remmers's General Scale	106
Application: Revising A Scale	108
Discussion	111
Cronbach's Alpha	111
Programs: SAS PROC Means, Alpha, r_{total} and SPSS	111

PART III: CLUSTERING **113**

Reverse Scoring for Negative Items	113
------------------------------------	-----

8. GRAPHIC SIMILARITY ANALYSIS **115**

Graphing Ability and Achievement	115
Graphing Letter Similarity	116
Graphic Analysis of Word Similarity	117
Elementary Linkage Analysis	118
Linkage Analysis of Test Scores	118
Discussion	119

9. SUCCESSIVE COMBINING **121**

Ward's Minimum Variance Method	121
Grouping Students on Reward Preference	124
CD-ROM and SAS Clustering Example	128
Discussion	131
Johnson's Nonmetric Single and Complete Link Clustering	132
Clustering the WISC Tests with HICLUS	134

10. PARTITIONING **137**

K-Means Iterative Clustering	137
Application: Visual or Auditory Preference for Reading Instruction	141
Discussion	142

11. HIERARCHICAL DIVISIVE **143**

Successive Splitting	143
Dividing By Largest Variance	143

Application: Grouping Ham Radios	144
Number Of Clusters	145
Graphing The Clusters	145

PART IV: MULTIDIMENSIONAL METHODS **147**

12. FACTOR ANALYSIS **149**

Representation of the Correlation Matrix	149
Trial and Error	151
Test Score Assumptions	152
Accountable Variance	153
Principal Components Analysis (PCA)	155
Factor Rotation	157
Specific Problems Associated With Factor Analysis	158

13. MAPPING INDIVIDUAL PREFERENCE **161**

Singular Value Decomposition	161
Carroll and Chang's Multidimensional Vector Model	162
MDPREF	164
CD-ROM Example Using MDPREF	165
Application: Occupational Ranking by Japanese	170
Inclusion of the Ideal Point	174
Ideal Point Projection	174

14. MULTIDIMENSIONAL SCALING **175**

How Kruskal's Method Works	176
SAS Analysis of Trevally Data	179
Application: Word Similarity (SAS MDS Using PEROVER Data)	180

15. INDIVIDUAL DIFFERENCES SCALING **185**

Output from INDMAT	185
SINDSCAL	185
CD-ROM Example of SINDSCAL With Learning Disability Data	186
How SINDSCAL Works	190
ALSCAL	190
Example with Dessert Data Using SAS Market	191
How ALSCAL Works	194
Alternating Search Analogy	195
Application: The Letter Wheel	196

APPENDIX A: Using a Computer to Solve Problems **199**

SAS 199

Format	200
--------	-----

Using the CD-ROM 200

- Readme General 202
- System Requirements 203
- Preparing to Run Programs 203
- Running the Programs 204
- Printing Reports 205
- Error Messages 205
- Troubleshooting 206
- Full File Names 206
- What is included on the CD-Rom for each program 207

Using the Internet 208

- Bell-Labs Netlib 208
- PC-MDS 209
- VISa 209
- The Three Mode Company 209
- ProGAMMA 209
- Scaling Methods and Terms 209

APPENDIX B: Tables 211

- Table A: Balanced Orders for Paired Comparisons for the Numbers from Five to Seventeen 212
- Table B: Selected Balanced Incomplete Block Designs 214
- Table C: Percentage Points of the Studentized Range for Infinite Degrees of Freedom 217
- Table D: Selected Range Values in the Two-Way Classification 218
- Table E: Cumulative Probability Distribution for Circular Triads Upper and Lower 10% Tails Across 5--15 Objects 219

REFERENCES 221

AUTHOR INDEX 229

SUBJECT INDEX 233

MAP OF SCALING METHODOLOGY 239

PREFACE

This text is written for instructors, students and researchers in the social and behavioral sciences who wish to analyze data that result from subjective responses. This edition concentrates on simplifying ways to handle data as opposed to the finer mathematics of how each program works and is addressed to general practitioners interested in the measurement and representation of attitudes. The methods presented have been chosen because they: (1) will handle the majority of data analysis problems; (2) are useful; (3) are easy to comprehend; and (4) have functional software solutions.

The second edition of *Scaling Methods* is prompted by the demonstrated value of the first edition in helping faculty and students to do research in the behavioral sciences. The senior author has taught a course in scaling methodology at the University of Hawai'i for three decades. *Scaling Methods* (Dunn-Rankin, 1983) has been the primary text for this course since the books' initial publication. During the last ten years the Department of Educational Psychology at the University of Hawai'i graduated 45 Ph.D. candidates. Over 40 percent of these graduates utilized some form of scaling methodology as part of their dissertation research. No other single course has had as much influence on the exploratory research of the department's students. In addition, a great many other doctoral candidates in such diverse fields as Communication and Information Sciences, Zoology, Library Science, Linguistics, Teaching English as a Second Language, Social Work, Public Health, Psychology, and Educational Administration have utilized scaling techniques in their dissertations.

What's New ?

The new text emphasizes **functionality**. The first edition was written to bring scaling methodology into use in behavioral science research. Unfortunately the necessary software existed mainly on paper in the back of the text or was isolated in places like Bell Laboratories or in a few University computer systems. Functional auxiliary FORTRAN programs were only available at the University of Hawaii or Florida State University, places where the senior author taught a course in scaling methods.

Professor Susan Wallace has converted these auxiliary programs and several other methodological programs to run on any personal computer with Windows. She has annotated

each of these programs and provided elaborate *readme files* and *error messages*. A supplement to this second edition is a CD-ROM of programs necessary to get raw data into matrix form such as the program, PEROVER. This specific software takes free clustering results and produces a percent overlap matrix. The CD-ROM also includes nine programs for unidimensional and multidimensional analysis. It includes programs such as TRICIR, and MDPREF. TRICIR does a circular triad analysis of paired comparison data. MDPREF takes the results of TRICIR and produces a multidimensional preference analysis. Details on how to use the CD-ROM are provided in Appendix A. One hundred pages of FORTRAN computer programs in the first edition have been deleted and most of these programs are now included on the CD-ROM.

In this second edition the authors have **simplified** the measurement process. The text illustrates that there are only four different kinds of tasks that one can ask of respondents. They are ordering, categorical ratings, free clustering and similarity judgments. Each of these tasks, such as paired estimates of similarity (judgments of similarity), lead to a specific matrix of responses. The matrix can then be analyzed by Multidimensional Preference Analysis (MDPREF). Different tasks require different types of responses and their analyses are not the same. To this end the authors have created a Map of Scaling Methodology (see page 238). This is a flow chart with the task types at the top, intermediate steps in the middle and final analyses near the bottom. This map is a micro representation of the text and the authors are pleased that it graces the cover of the book. After every task presented in the introduction a specific flow chart is included.

The basic methods and statistical measures presented have not radically changed. But their presentation has been **rewritten**. Professor Shiqiang Zhang has gone over every calculation in the text and corrected specific errors, errors of omission and statistical errors pointed out by the reviewers. A new chapter on Order Analysis is introduced. The chapter on Factor Analysis has been extensively rewritten. Every chapter has been rewritten and many new examples included. For example, the introduction now includes a simple illustration of a two dimensional solution which can be solved with paper and pencil.

The authors have **changed the format**. The second edition is larger in overall size and the font size is larger. A more detailed table of contents with extensive page numbering is included. **Reference** is made to the extensive and sophisticated methodologies available in the literature. The authors have included in the text the methods thought to be most functional. SAS and SPSS have very complete software programs for statistical analysis. The text identifies with SAS and has included examples of the set up and results for its programs such as principal components analysis, MDS and Proc Corr for calculating correlations.

More sophisticated software exists on the Internet and the World Wide Web has made scaling methodology more available than ever. For those educators who are interested, the authors have included in Appendix A some information on using the Internet to do analyses not included in the text. For example, Thurstone's Case V, is described in the text but Tucker and Gulliksen's software, while printed in the first edition is not extant. Readers, however, can obtain for a fee a program called Case5 from *marketing.byu.edu*.

The overall goal of the new edition is to make scaling analysis more functional. A course in scaling methodology, using a draft of the new text, has been taught for three summers at the University of North Texas in Denton. The results were positive. The use of the CD-ROM was

easily assimilated by the students and the application of various methods to their particular research was useful and productive. Students using the text should be familiar with computers and an initial course in statistics would be helpful but not necessary.

Content and Organization

Part I of the text introduces the major purposes for the analysis of psychological objects, particularly any variable for which some degree of attitude or perception can be measured. Scaling analyses attempt to produce estimates of the distance between each pair of psychological objects and to provide a parsimonious representation of the objects; that is, a simplified picture (a map of the data). Such objectives can have useful consequences such as enhancing the validity of attitude measuring instruments to the discovery of new relationships underlying a set of objects. Part I also provides an introduction to the types of tasks that an experimenter can initiate. It details the major ways in which measures of proximity (similarity or distance) can be obtained from responses to the objects of interest.

The next three parts of the book explain the various methodologies. A gradual progression from simple representations to more complex is presented. The methods start with Part II: Unidimensional Techniques, move to Part III: Clustering, and end with Part IV: Multidimensional Analyses. The authors have found that if students learn the early techniques first, the latter methods are more easily assimilated.

The text follows the pedagogy of instruction by example. Each chapter presents (1) an exposition of the theory surrounding the particular methodology; (2) a simple example; (3) real-world application(s); and (4) references to a computer solution. Each method is a complete unit, so readers may turn directly to the chapter that explains a specific methodology.

Acknowledgements

The authors are indebted to Rebecca Swartz for her initial editing and contribution of the ORDER program to the CD-ROM. Pat Dunn-Rankin's many suggestions and editing help are greatly appreciated. Michael Gallia, Rhonda Christensen, Cesar Morales and many other readers have assisted in reviewing the manuscript as well. The authors are indebted to Layne Wallace for the development of the Scaling Methods website at <http://www.iittl.unt.edu/scaling>. The authors are indebted to the reviewers of the initial manuscript John C. Caruso, University of Miami, Shlomo S. Sawilowsky, Wayne State University and Xiang Bo Wang, The College Board.

This page intentionally left blank

PART I

FOUNDATIONS

The foundations of scaling methods contain:

1. the definition of relative measurement,
2. the kinds of instruments or tasks that can be responded to, and
3. the measures of proximity that can be applied to the responses that have been gathered.

This page intentionally left blank

1

SCALING DEFINED

Relative Measurement

The Fahrenheit Scale

Scaling consists of measuring and comparing objects in some meaningful way. The process includes some visual representation, usually a linear or multidimensional map. A thermometer is an example of a linear representation. One cold winter, Gabriel D. Fahrenheit surrounded a glass tube, containing mercury, with a mixture of snow and salt. He made a mark on the tube at the height of the mercury and called this point zero. He knew that if the mercury ever went that low again it would be very cold. He had, in fact, attached significant, if relative, meaning to the height of the mercury.

The mercury heights for freezing and boiling water were also indicated on the tube. The distance between the freezing and boiling marks was divided into 180 equal parts or units. The snow-salt mark was observed to be 32 of these units below the freezing point of water. Thus the freezing point of water was given as 32°F or 32 degrees on the Fahrenheit scale and the boiling point became 212°F. Fahrenheit had created a relative scale for assigning temperatures to mercury heights.

Psychological Objects

In the social sciences, researchers are continually trying to measure and compare human perceptions. They (a) create scales by assigning psychological objects to numbers and then (b) locate individuals on the scale they have created. Psychological objects can be tangible, such as cars and postcards, but they can also be anything which is perceived by the senses resulting in some attitudinal response. Psychological objects can be colors, words, tones, and sentences as well as houses, gold stars, and names or pictures of television stars. Psychological objects are most often presented as sentences or statements such as "There will always be wars" or "I hate war." With young children, the objects are often pictures.

As an example, look at the following scale on attitudes toward reading

- 6 — When I become interested in something, I read a book about it.
- 5 — I almost always have something I can read.
- 4 — I read when there is nothing else to do.
- 3 — I only read things that are easy.
- 2 — I never read unless I have to.
- 1 — I seldom read anything.

Although there are several ways to score or place individuals, one way is to ask a respondent to indicate which sentence best describes her or his attitude toward reading. Different people might choose different answers. Respondents would then be placed at different positions on the scale. Although this scale is short and reading interests are rarely in just one dimension, the scale can differentiate subjects with varying reading interests.

Mapping

It is a basic problem of scaling to determine the proximities (similarities or distances) between a set of objects and then locate or map these objects onto the smallest space that will effectively retain the basic information about the data. This projection may reveal the underlying structure or unique relationships among the items. Subsequently, it can provide relative positions of individuals with regard to the mapped stimuli.

Introduction to Scaling

Suppose one is interested in the special education problem of mainstreaming children with disabilities into the regular classroom. Four disabilities are chosen. The psychological objects are:

- (LD) Learning Disabled
- (MR) Mentally Retarded
- (D) Deaf (hearing impaired)
- (B) Blind (visually impaired).

The disabilities are paired in the six $[K(K-1)/2]$ possible ways, where $K = 4$ is the number of disabilities. They are presented in the following task:

Pairings		Similarity
LD	MR	_____
LD	D	_____
LD	B	_____
MR	D	_____
MR	B	_____
D	B	_____

First, teachers are asked to **judge** the similarity between two members of each pair of disabilities using a number between 1 and 10. A 10 indicates that the pair is very similar and a 1 indicates the pair is very dissimilar. This value is written in the blank to the right of each pair. Second, the teachers are asked to **choose**, by marking with a + the type of disability student, in each pair, they prefer to teach in a regular classroom. The responses for five teachers are presented below.

T1 Pairings		Similarity	T2 Pairings		Similarity	T3 Pairings		Similarity
+ LD	MR	<u> 4 </u>	+ LD	MR	<u> 6 </u>	+ LD	MR	<u> 5 </u>
+ LD	D	<u> 4 </u>	+ LD	D	<u> 3 </u>	+ LD	D	<u> 3 </u>
+ LD	B	<u> 4 </u>	+ LD	B	<u> 2 </u>	+ LD	B	<u> 4 </u>
MR	+ D	<u> 5 </u>	MR	+ D	<u> 8 </u>	+ MR	D	<u> 6 </u>
MR	+ B	<u> 2 </u>	MR	+ B	<u> 2 </u>	MR	+ B	<u> 3 </u>
D	+ B	<u> 5 </u>	D	+ B	<u> 4 </u>	D	+ B	<u> 4 </u>

T4 Pairings		Similarity	T5 Pairings		Similarity
+ LD	MR	<u> 2 </u>	+ LD	MR	<u> 2 </u>
+ LD	D	<u> 2 </u>	+ LD	D	<u> 6 </u>
LD	+ B	<u> 6 </u>	+ LD	B	<u> 6 </u>
MR	+ D	<u> 4 </u>	+ MR	D	<u> 7 </u>
MR	+ B	<u> 2 </u>	MR	+ B	<u> 4 </u>
D	+ B	<u> 5 </u>	+ D	B	<u> 5 </u>

The medians of the five teacher similarity data are quickly determined by ordering each teacher's ratings and selecting the middle value. These similarity values are placed in a similarity matrix as follows:

Median Similarities for Disabilities				
	LD	MR	D	B
LD				
MR	4			
D	4	6		
B	6	2	5	

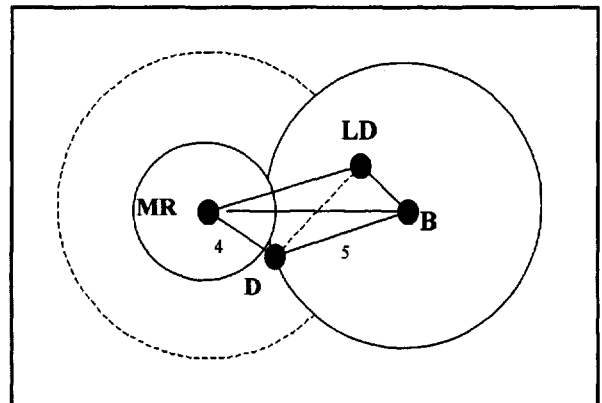
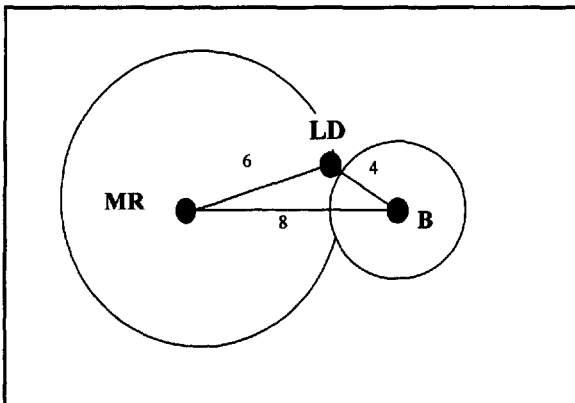
In order to map the data, the similarities are subtracted from 10 (maximum possible similarity). This creates a matrix of distances.

Distances = 10 – Similarity

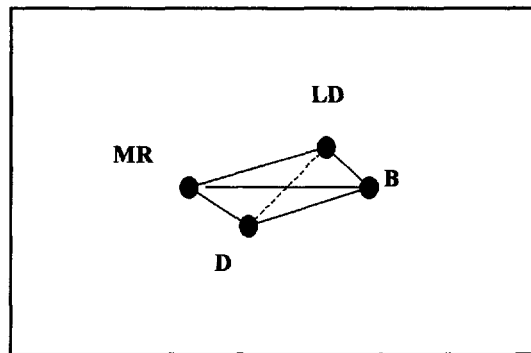
	LD	MR	D	B
Learning Disabled				
Mentally Retarded	6			
Deaf	6	4		
Blind	4	8	5	

A map is created by drawing circles using the distances as radii. First, the longest distance in the matrix (8) is between MR and B. A line is drawn 8 units long. Using MR as a center, a circle with a radius of 6 is drawn. This is the distance between MR and LD. Now, using B as a center, a circle of radius 4 is drawn (the distance between B and LD). Where the two circles intersect is the relative position of LD (see figure on the left) below.

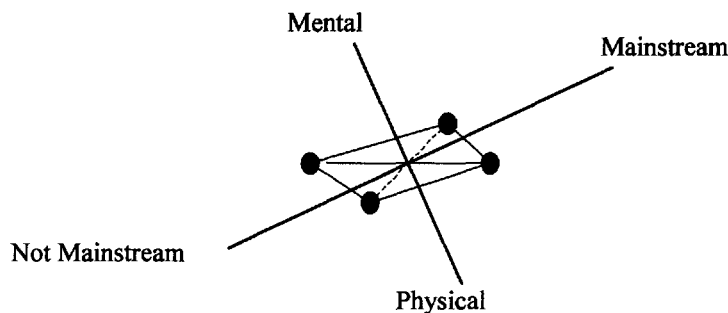
Next, using MR as a center and 4 (the distance between MR and D) as a radius, a new circle is drawn. Using B as a Center and a radius of 6, another circle is drawn fixing the position of D. Hopefully, the remaining distance between LD and D is close to the tabled value. In this case, a small compromise is made to accomodate all the distances by moving D slightly farther away from MR and LD and closer to B.



The final mapping shows the relative positions of the disabilities. It suggests that there are relative similarities and indicates that two dimensions can adequately accommodate the data.



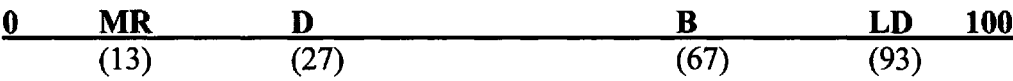
Meaning can be attached to these two dimensions based on the original direction of the research. The placement of the configuration of the disabilities in the space is arbitrary and as long as the distances are maintained the objects can be rotated. In this case, arbitrary axes can be drawn through the figure and dimensions assigned. Learning Disabled and Blind are suggested candidates for regular class instruction. Deaf and Blind are physical disabilities and they are contrasted with Mental disabilities.



The **preference votes** are tallied by counting the plus (+) signs indicating preferences for each disability by each respondent. The vote vectors are tabled and summed below:

Mainstreaming Votes				
	LD	MR	D	B
R1	3	0	1	2
R2	3	0	1	2
R3	3	1	0	2
R4	2	0	1	3
R5	3	1	1	1
Total	14	2	4	10

When each total is divided by 15 (the maximum number of possible votes) and multiplied by 100 the results can be displayed on a line as shown below.



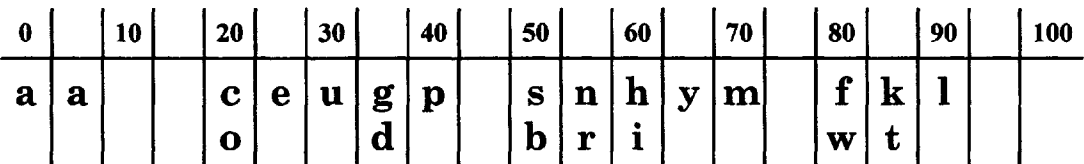
(e.g. for LD, 14 votes/15 = .93, .93 X 100 = 93)

The unidimensional map illustrates that Learning Disabilities and Blind are preferred as mainstreaming candidates by the five teachers.

This exercise is used as a conceptual introduction to scaling. Hand-calculated methods can be applied to various stimuli of interest. But, increasing the number of objects, the number of judges and utilizing a variety of tasks makes hand solutions difficult. This text will illustrate ways to make larger problems more tractable. The reader may compare this result with a sophisticated analysis, TRICIR Analysis Summary (p. 71) and with SINDSCAL, (p. 186).

Euclidean Space

Generally, Euclidian space provides a framework within which numbers can be assigned to objects in a relative but meaningful way. The use of one-dimensional space is demonstrated by the scaling of lowercase letters of the English alphabet (letters are the psychological objects) on a unidimensional or linear scale. Twenty-one letters are arranged in terms of their similarity to specific target letters (see Fig. 4.4 p. 63). Note that when the letter **a** is used as a target, the other letters are scaled in their perceived similarity to **a** as follows:



In this scale, the letter **l** is seen as least similar to the target letter **a**; whereas **c**, **o**, and **e** are judged to be much closer to **a**. Relative meaning can be attached to the ends of the scale (i.e., the numbers 0 and 100). They represent the unlikely prospect of having every one of the judges (315 second- and third-grade children) indicate that the same letter was **most** like the target letter **a** and that one other letter was **least** like the target letter **a**, when presented with all possible pairs of 21 letters. In this scale, the distance between **a** and **c** is shorter than the distance between **a** and **y**. One can infer from the scaling technique that **y** is more different from **a** than **c**.

Multidimensional space (2, 3, or more dimensions) is used to display or map the distances between objects that cannot be effectively placed on a linear scale.

Guttman Scales

Distances are not a necessary prerequisite for a scale. One could select a set of objects for which its order is the scale. If, for example, the following math problems, ordered in difficulty, were presented to a group of school children it would be known as a Guttman scale.

- (1) $2 + 2 =$
- (2) $24 - 16 =$
- (3) $375/4 =$
- (4) $4!/2!(4 - 2)! =$
- (5) $d(3x^2 + 4) / dx =$

Each succeeding problem is more difficult than the one before it. The questions or psychological objects constitute a scale based on difficulty. If we score a 1 for each correct answer and 0 for an incorrect answer, the pattern of ones and zeros over the five questions tells us where the student is on this math difficulty scale. Thus a person who has the pattern 11110 is farther along on the scale than the student with scored responses of 11000. In a perfect scale, a single number (the **sum** of the correct responses, 4 vs. 2) determines where a student is with regard to such items. Such scales are called *deterministic*. Responses such as 01111 might indicate random error or challenge the ordinal properties of the items.

Judgments or Choices

Figure 1.1 presents an outline for attitudinal measurement. First, the direction or focus of the instrument is defined. Then psychological objects, selected or created, are presented in a task. If the task requires **judgments**, methods are used to assign meaningful numbers to the objects. From such an analysis, a subset or subsets of the objects are chosen and formulated into a scaling instrument. This instrument can then be presented to the target group(s) and the responses scored. When **choices** are obtained, a descriptive analysis occurs directly. Such analyses can generate or test hypotheses.

Judgments are objective ratings of similarity, order, or value. It is possible despite a particular bias to act as a judge. One can rate "I like school" as a more positive statement than "School is OK" despite how he or she feels.

Choices or preferences are subjective and should reflect a personal point of view. In the example given in the introduction to scaling, both judgments and choices were obtained on the same instrument.

Generally, however, judgments of the similarity between the psychological objects are obtained initially. This can be done using paired comparisons, free clustering, etc. (see Similarity Judgments, p. 27). When the structure of the instrument, based on judgments is determined, then it is used as a scale. The iterative nature of the scaling process is suggested in Figure 1.1. It is generally important to judge, analyze the judgments and reformulate the task or instrument before its final administration to insure its validity and reliability.

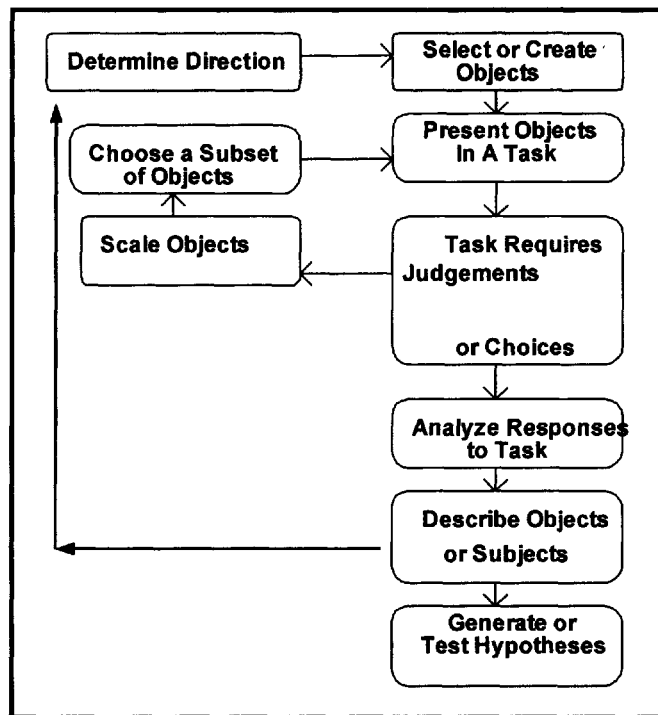


FIG. 1.1. The circular nature of the scaling process.

The first step in scaling, **Determining Direction**, is based on the theoretical rationale surrounding the area of interest. A certain amount of argument is necessary to provide a framework for selecting or creating the psychological objects. Tasks can vary widely and some are easier and less valid than others. Once a trial instrument is developed, choices are analyzed and a description may occur or the direction may be changed.

2

A limited number of tasks in educational and psychological scaling have been established. The tasks can be divided into four areas. These include: (1) **Ordering Tasks**, (2) **Categorical Ratings**, (3) **Similarity Judgments**, and (4) **Free Clustering** (see Table 2.1).

Tasks for Assessing People's Judgments or Choices About Psychological Objects

Tasks	Examples
Ordering	Who or What is best, next best, etc.
Categorical Rating	Onions: Good : __:__:__:__:__: Bad
Similarity Judgments	How similar are fish and chicken? __
Free Clustering	Put the words that are similar together.

Differences in these primary tasks create differences in the direction and kind of analyses that can be performed on the resulting data. Generally, however, a similarity matrix of some kind forms the basic data set. Similarities are measures, like correlations, in which the larger the index the more similar are the two objects being compared. A matrix of similarities is a collection of similarity values for all pairs of objects. Similarities are changed into distances in order to view the objects more effectively. This is often done by subtracting similarities from a constant. The constant is usually the largest value in the set of similarities.

Ordering

Ordinal tasks involve ranking psychological objects in some way to produce dominance data (Shepard 1972a), that is, one stimulus dominates another. Such data are often called nonmetric because only judgments of greater than ($>$) or less than ($<$) are required. If we ask a class to line up according to height (shortest to tallest) this is a ranking or ordering task.

Paired Comparisons

Ranking can be accomplished directly or derived from pairing the objects and counting the votes for each pair. The votes are inversely related to a ranking and can be called rank values. For example, three statements: (a) Teacher gives you an A, (b) Friends ask you to sit with them, and (c) Be the first to finish your work, are paired in all possible ways as shown below:

- ♦ (a) Teacher gives you an A.
 (b) Friends ask you to sit with them.
- ♦ (a) Teacher gives you an A.
 (c) Be the first to finish your work.
- ♦ (b) Friends ask you to sit with them.
 (c) Be the first to finish your work.

In this example, the (♦) represents the choice of a particular statement in each of the pairs. By counting the votes for each statement, it is determined that (a) gets 2 votes, (b) 1 vote and (c) 0 votes and the rank values for these statements allows us to establish the rank order. Rank order usually assigns a (1) to the best or highest rank. Votes are, therefore, a direct but inverse reflection of typical ranks.

Circular Triads

It is possible for the votes to be circular. That is, a subject may like (a) better than (b), (b) better than (c), but (c) better than (a). This results in tied votes. The analysis of circular triads is an interesting addition to establishing the scale values of objects which have been paired and voted upon (Knezek, Wallace, & Dunn-Rankin, 1998).

Partial Ranks and Balanced Incomplete Block Designs

Gulliksen & Tucker (1961) illustrated a compromise between direct ranking and complete paired comparisons. This scheme involves the use of Balanced Incomplete Block (BIB) designs. When the number of objects becomes larger than 20 (20 objects involves $K(K-1)/2 = (20)(19)/2$ or 190 pairs), then the time needed for a subject to vote on **all** the pairs becomes increasingly tedious. In BIB designs, small subsets of the objects or statements are grouped in such a way that all possible paired comparisons can be inferred by directly ranking the objects in each small subset (Gulliksen & Tucker, 1961). One of the simplest of the BIB designs involves seven subsets of three objects. This design, sometimes called a Youden Square, is presented on the right. In such arrangements, each object is compared to each other object just once. The 21 pairwise comparisons have been collapsed into 7 simple rankings. Table B in Appendix B shows six designs for various numbers of Objects. Cochran & Cox (1957) provide many others.

Youden Square

a_	b_	d_
b_	c_	e_
c_	d_	f_
d_	e_	g_
e_	f_	a_
f_	g_	b_
g_	a_	c_

In the Youden Square, the simplest BIB design, the task is to rank order three objects at a time. The objects are related to the letters in each block or row of the design. Suppose the objects were adjectives and a single subject ranked the adjectives in **each** row in terms of which he or she would most like to be: the smaller the value, the more the person wants to have that particular characteristic.

(a) powerful_3_	(b) rich_1_	(d) good-looking_2_
(b) rich_2_	(c) honest_1_	(e) generous_3_
(c) honest_1_	(d) good-looking_3_	(f) famous_2_
(d) good-looking_2_	(e) generous_3_	(g) intelligent_1_
(e) generous_2_	(f) famous_1_	(a) powerful_3_
(f) famous_2_	(g) intelligent_1_	(b) rich_3_
(g) intelligent_1_	(a) powerful_3_	(c) honest_2_

Then rank values (votes) are derived by establishing a matrix in which a "1" is inserted if the column object (adjective) is judged or preferred over the row object. For example (g) intelligent is preferred over all other characteristics and its column sum is 6.

	rich	honest	good-looking	generous	famous	intelligent	powerful
	b	c	d	e	f	g	a
b		1			1	1	
c						1	
d	1	1			1	1	
e	1	1	1		1	1	
f		1				1	
g							
a	1	1	1	1	1	1	
Sums	3	5	2	1	4	6	0

The sum of the votes can then be utilized as a profile of ordered data for a given subject. No missing votes are allowed. Although this simple illustration can be analyzed by hand, a computer program is needed to convert the data for larger designs. The CD-ROM that accompanies this text contains a program which converts **BIB** data to rank preference profiles. Computer output for the example above is listed under student 2 on page 17.

Direct Ranking

Direct ranking consists of assigning integers to objects, indicating order of preferences or judgments. If there are k objects the integers will run from 1 to k .

Example: Given the following occupations, participants are asked to order them on the basis of their personal preference. In this case a rank value of 1 is given the most desired occupation.

Occupations	Occupations Ranked					
	Judge	C	R	M	P	F
Carpenter						
Ranger	1	5	3	4	2	1
Mailworker	2	3	4	2	5	1
Police Officer	3	4	5	3	2	1
Firefighter	4	4	5	3	1	2
	5	3	5	2	4	1
	Σ Ranks	19	22	14	14	6
	Σ Votes	6	3	11	11	19

Ranks and Rank Values

If respondents are to rank more than a few objects, initially splitting the objects into two groups of the most and least desired occupations is a good first step. Providing room, on paper, for sorting the ranks is also necessary. Ranks usually run from 1 to k with 1 given to the most desired object and k the least. Because the votes in paired comparison studies are an inverse of ranks, they are called rank values. Votes can provide a positive scale of interest for each occupation. Votes = $(k - R_k)$ where k = the number of objects and R_k is the rank of object k . For the direct inverse use $(k + 1) - R_k$.

Tetrads (Pairs of Pairs)

Pairs of pairs may be created by the experimenter. Using the four objects man, woman, boy, and girl generates six pairs and this in turn generates $(6)(5)/2 = 15$ pairs of pairs. For example:

Pairs of Pairs		
man-woman -- man-boy	man-boy -- man-girl	man-girl -- woman-girl
man-woman -- man-girl	man-boy -- woman-boy	man-girl -- boy-girl
man-woman -- woman-boy	man-boy -- woman-girl	woman-boy -- woman-girl
man-woman -- woman-girl	man-boy -- boy-girl	woman-boy -- boy-girl
man-woman -- boy-girl	man-girl -- woman-boy	woman-girl -- boy-girl

The 15 "pairs of pairs" can be handled like other paired data. Participants, for example, might be asked which of each pair of pairs is closer or more similar. One can also ask the judge to indicate degrees of similarity. Such data are useful in looking at the dimensions of **relationships**, within a family, for example. Because the pairs of pairs increase dramatically with the number of objects, this may restrict the use of pairs of pairs. For example, 11 objects generate 1,485 pairs of pairs.

Arranging Pairs

It has been customary to arrange the objects in pairs according to the method outlined by Ross (1934). Table A in Appendix B presents balanced orders for the presentation of pairs for odd numbers of objects from 5-17. For even numbers, the next higher odd set of pairs is used, striking out all pairs containing the nonexistent odd object. Ross' pairing for five objects is as follows:

$K = 5$ 1-2, 5-3, 4-1, 3-2, 4-5, 1-3, 2-4, 5-1, 3-4, 2-5

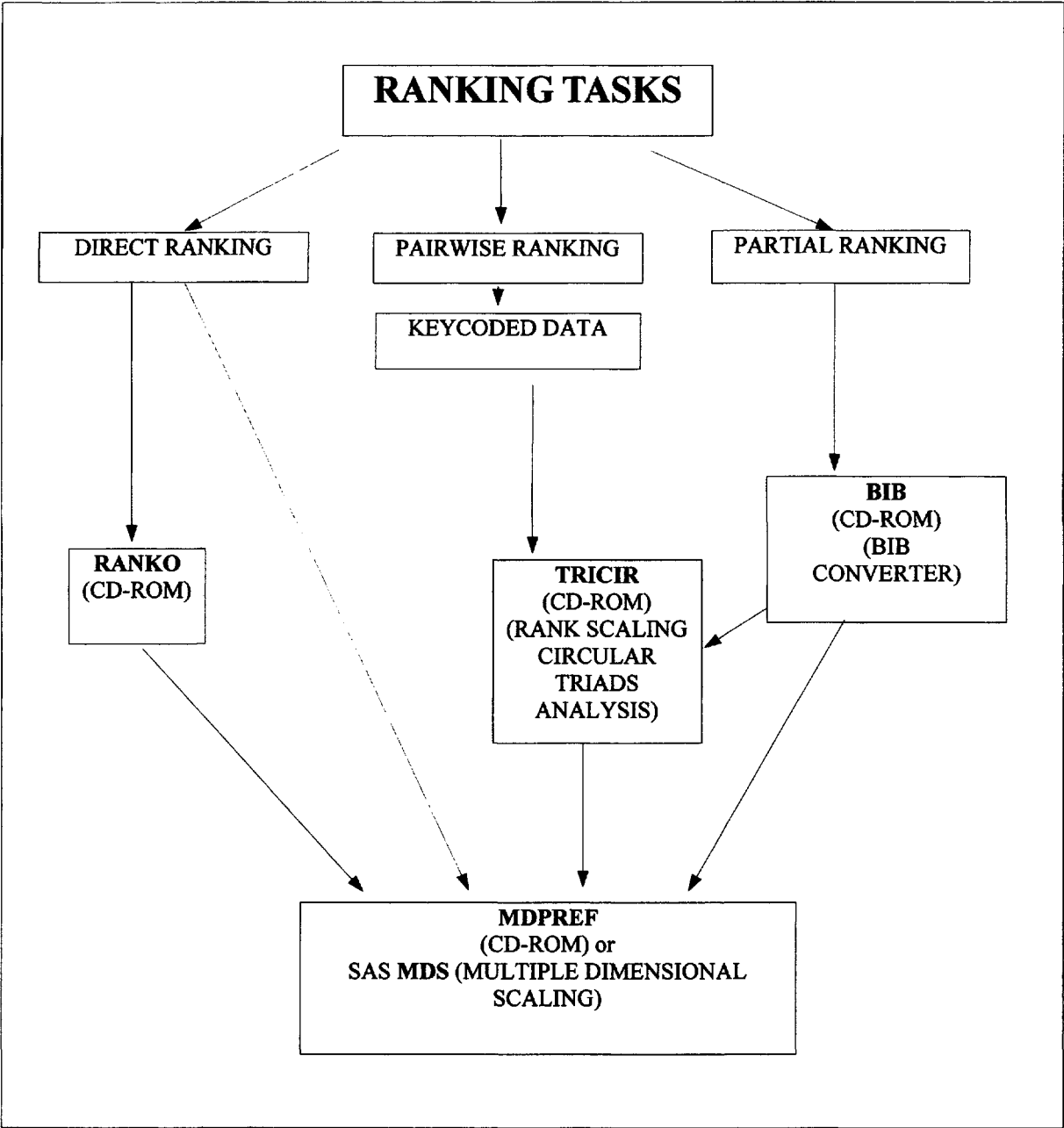
Pair arrangements may be randomized if care is taken to randomize both the order of the pairs and positions of the objects in the pairs. For example:

Not random	$K = 4$ 1-2, 1-3, 1-4, 2-3, 2-4, 3-4
Random	$K = 4$ 3-2, 2-4, 2-1, 4-3, 3-1, 1-4

Flow Diagram for Analysis of Ordinal Tasks

An object's relative position resulting from Direct Ranking or from Pairwise Ranking can be determined by averaging the rank values of the objects and testing them for statistically significant differences (See **RANKO** on the CD-ROM). **TRICIR** (for circular triads) is a program on the CD-ROM that scales the data and tests the data for circular judgments. There is also a program for Complete Paired Comparisons, **COMPPC**, that handles pairs under Thurstone's Case V normality assumptions. The FORTRAN program is found in Dunn-Rankin, (1983).

When Balanced Incomplete Block Designs (BIB) are used to create partial ranks, a program (**BIB**) on the CD-ROM is utilized to convert the partial ranks into one vector of the object's rank values for each subject. (See BIB Example, p. 17.) In this example, responses to the instrument on page 13 by two participants are converted into rank profiles. In order to utilize the CD-ROM refer to Appendix A. The vector of integers can be utilized as input into Multiple Dimensional Preference Analysis **MDPREF** which places the profile of the individual in the space of the objects indicating their dimensional preference. This program is also on the CD-ROM. The following figure illustrates the flow of the analyses. For sophisticated users SAS Market also does such analyses.



BIB Example on CD-ROM

Configuration File (bib1.cfg)

Bib Program Title
2 7 7 3 1
bib1.dat
bib1.out

Title
2 subjects, 7 objects, 7 blocks, 3 per block, 1 full output
Input file
output file

Input File (bib1.dat)

7 1 3 1 2 4 2 3 5 3 4 6 4 5 7 5 6 1 6 7 2
1 3 2 3 1 2 1 3 2 1 3 2 3 1 2 1 2 3 3 1 2
3 1 2 2 1 3 1 3 2 2 3 1 2 1 3 2 1 3 1 3 2

BIB arrangement of objects and blocks
First Subject's Ranks
Second Subject's Ranks

Output File (bib1.out)

Bib Program Title
NUMBER OF OBJECTS = 7
NUMBER OF BLOCKS = 7
OBJECTS PER BLOCK = 3
NUMBER OF SUBJECTS = 2

7 1 3
1 2 4
2 3 5
3 4 6
4 5 7
5 6 1
6 7 2

Student 1
1 3 2
3 1 2
1 3 2
1 3 2
3 1 2
1 2 3
3 1 2

Student 2
3 1 2
2 1 3
1 3 2
2 3 1
2 1 3
2 1 3
1 3 2

1 2 3 4 5 6 7
1| 0 1 1 1 1 1 1
2| 2 0 2 2 2 2 1
3| 2 1 0 2 1 2 1
4| 2 1 1 0 1 1 1
5| 2 1 2 2 0 2 2
6| 2 1 1 2 1 0 1
7| 2 2 2 2 1 2 0
votes | 0 5 3 1 5 2 5

1 2 3 4 5 6 7
1| 0 1 2 2 1 1 2
2| 2 0 2 2 2 1 2
3| 1 1 0 2 1 1 2
4| 1 1 1 0 1 1 2
5| 2 1 2 2 0 1 2
6| 2 2 2 2 2 0 2
7| 1 1 1 1 1 1 0
votes | 3 5 2 1 4 6 0

The votes in the Profile Vectors make sense when you realize that the arrays are formed by having a 1 represent the choice of a column object over a row object and the 2 represents a non vote. The profile is found by summing the ones (1) in each column.

Categorical Ratings

Ordered Categories subsume many of the most frequently utilized unidimensional scaling tasks. An example is provided in Fig. 2.1. These measures are commonly referred to as Summated Ratings, Likert Scales, or Successive Categories. Such titles, however, refer to different assumptions about the data and different analyses applied to the ordered categories data rather than to the task itself.

For the following statements A = Agree, TA = Tend to Agree, TD = Tend to Disagree, and D = Disagree. Mark an X in the appropriate box.				
	A	TA	TD	D
1. I will be lucky to make a B in this class.	: _ : _	: _ : _	: _ : _	: _ : _
2. This class has a tough professor.	: _ : _	: _ : _	: _ : _	: _ : _
3. This is the kind of class I like.	: _ : _	: _ : _	: _ : _	: _ : _
4. I would not take this class if it wasn't required.	: _ : _	: _ : _	: _ : _	: _ : _
5. The demands for this class will not be high.	: _ : _	: _ : _	: _ : _	: _ : _

FIG. 2.1. The general form for an ordered category task.

Likert (1932) suggested that statements (psychological objects) should be written so that people with different points of view will respond differently. He recommended that statements of similar content vary widely in emphasis. For example, the statements "I would recommend this course to a friend," and "This is the worst course I have ever taken" will evoke different responses, but are generally evaluative in nature or dimensionality .

Specifically, social scientists should use statements that:

- (a) refer to the present or future rather than the past;
- (b) have only one interpretation;
- (c) are related to the continuum under consideration;
- (d) will have varied endorsements;
- (e) cover the range of content;
- (f) are clear;
- (g) are direct;
- (h) contain simple words;
- (i) are short (rarely over 20 words);
- (j) contain a complete thought;
- (k) are not universal,
- (l) are usually positive, and
- (m) avoid double negatives.

Judgments

When creating a scale, asking for judgments should be an initial step. Respondents are asked to determine a degree of similarity about the objects. With categorical ratings judging is a rare occurrence because the form lends itself to making choices, that is, indicating a preference.

Judgments can be determined, however, by creating category designations as degrees of judgment rather than preference. In this case, categories can vary from Very Favorable to Very Unfavorable. A survey might initially ask special education teachers to judge the severity of behaviors in an ordered category task that runs from (N) Normal, (SA) Somewhat Abnormal, (A) Abnormal, to (VA) Very Abnormal. For example:

	N	SA	A	VA
1. Threatening a teacher.	: ____ :	: ____ :	: ____ :	: ____ :
2. Spitting on the floor.	: ____ :	: ____ :	: ____ :	: ____ :
3. Chewing gum in class.	: ____ :	: ____ :	: ____ :	: ____ :
4. Fighting another student.	: ____ :	: ____ :	: ____ :	: ____ :

The Semantic Differential

The Semantic Differential (Osgood, Suci, & Tannebaum, 1957) has pairs of adjectives that anchor a categorically ordered scale. All of the adjectives are related to a central psychological object or concept. For example:

poetry									
valuable	_____	:	_____	:	_____	:	_____	:	worthless
good	_____	:	_____	:	_____	:	_____	:	bad
interesting	_____	:	_____	:	_____	:	_____	:	boring
easy	_____	:	_____	:	_____	:	_____	:	hard
light	_____	:	_____	:	_____	:	_____	:	heavy
simple	_____	:	_____	:	_____	:	_____	:	complex

Intercorrelations among the bipolar pairs of adjectives are determined after the appropriate administration. This matrix of similarities is then analyzed in order to find subsets of similar adjective pairs that constitute distinct dimensions in attitudes towards the concepts or objects.

Simple Scoring

Most common in the behavioral and social sciences are survey instruments of the type shown in Fig. 2.1 (p. 18). If k equals the number of categories, then the arithmetic **average** of the integers 1 to k assigned to the subjects' responses (checks, marks, or circles in the ordered categories) is often immediately utilized as a variable. This requires that the category names be carefully chosen in order to reflect equal weighting for each interval between them (where a weighting of one is applied). A unidimensional scale and unit weighting are assumed under Likert Scaling.

A program (TSCALE), first developed by Veldman (1967), is provided on the CD-ROM. The program orders each item in the instrument based on the frequency of all the subjects' categorized responses. Green's Successive Interval Scaling (1954) can be used to produce a unidimensional scale for summated ratings that controls for different interval widths. Scores can be the average of the responses to a single statement, a set of statements, or an average for all the statements in the survey.

Subsets of Items

It is popular to create from reflection or introspection an ordered category instrument and immediately evaluate the participants by averaging their responses to the items. This action violates a first step in scaling which is to search for **judgments** of similarity between the statements or psychological objects. Similarity judgments are necessary in order to substantiate the instruments construct and content validity. Because most instruments will likely be multidimensional, **subsets** of items representing latent traits, factors, or dimensions will generally be established. Traits, factors, and dimensions reflect the construct validity of the instrument that is being created. Their meaning is derived from matching the content of the subset to some theoretical framework, sometimes called a nomological net. That is, a network of arguments and research supporting the idea. For example, if a subset contained items such as:

It bothers me to see a grown man cry.
I don't like to look at an injured animal.
I don't like to look at pictures of poverty.
I feel bad when I see a dead cat in the road.

One might suggest an underlying dimension of empathy.

Subsets of items have the probability of being more valid and more reliable than single items. This is because reliability is positively related to the number of effective items and a collection of items is generally necessary to represent a construct effectively.

Steps in Ordered Category Scale Construction

It is effective to work on ordered category scales in a series of steps.

1. The items are created or obtained that have reasonable construct and content validity.
2. These items are **judged** in terms of their similarity, perhaps using the task of free clustering, and analyzed. Or the items may be administered directly to a reasonable sample of respondents (one that represents the target population).
3. The responses are coded into a subjects by items data matrix.
4. The items are interrelated using correlations or perhaps some measure of distance.
5. This produces a matrix of proximities which can be analyzed by **Cluster analysis**, **Multidimensional Scaling (MDS)** or **Factor Analysis**. These methods are used to determine outliers and collect effective subsets of items.
6. The subsets are analyzed using **Test** statistics calculating item difficulty, internal validity, and reliability
7. Then the instrument is reorganized and administered to the target population.

Ordered Category Example

The Ordered Category task may be modified in many ways. In the example below, circles have been used around the letters representing the category description as opposed to making checks on a line. In this case SA represents Strongly Agree. DS represents Disagree Strongly and so forth.

1. The instructor appears well organized.	Ⓐ	A	U	D	DS
2. The instructor is interested in the subject.	SA	A	Ⓐ	D	DS
3. This course has improved my cognitive skills.	SA	Ⓐ	U	D	DS
4. Morale in class has been positive.	SA	A	Ⓐ	D	DS
5. The instructor is sensitive to student feelings.	SA	Ⓐ	U	D	DS

Some researchers like to use only positive or negative adjectives such as Agree (A), Tend to Agree (TA), Tend to Disagree (TD), and Disagree (D) because this eliminates the neutral response.

The tendency for examinees to give only positive ratings with this type of instrument has had researchers utilize categorical descriptions such as (A) In the top 1% of all teachers; (B) In the top 5% of all teachers but less than 1%; (C) Among the top 25% of all teachers but less than; etc.

Restrictions of Ordered Categories

Ordered Category tasks result in a similarity matrix (correlations) between all the items. Many participants responses are required before the relationships between the items can be obtained. A single person's responses are insufficient for any individual differences analysis.

Number of and Naming of Categories

How many categories should an instrument display? The consensus is from 3 to 9 categories with the odd numbers 5 and 7 most preferred. For young children, 3 categories may be all they can handle and 9 categories requires a lot of thought and response time.

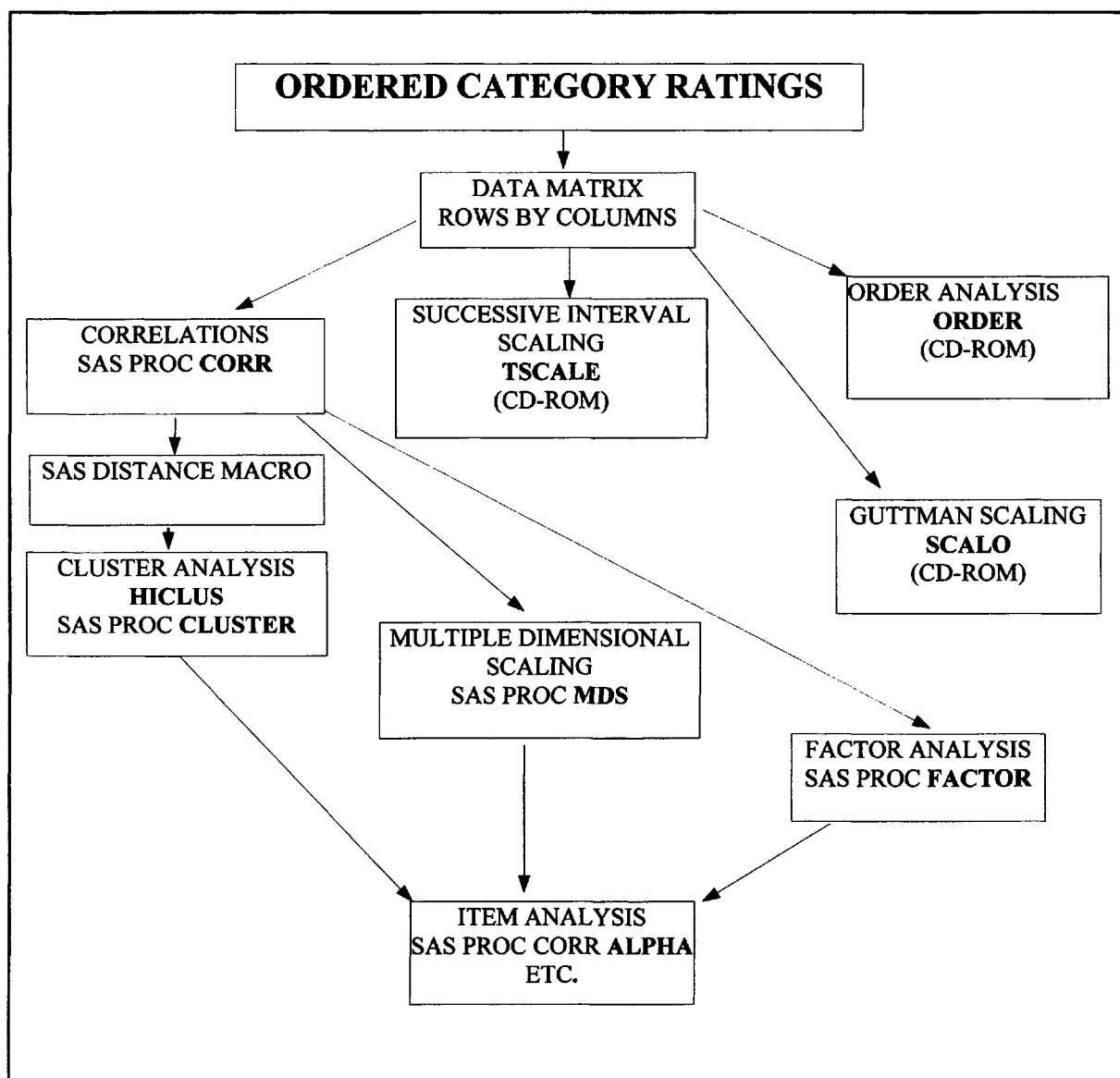
What should the categories be named? Zhang (1995) studied 21 relevant scales and deduced a stable semantic model of seven adverbs in the form of the following rank order:

- 1 extremely
- 2 unusually
- 3 decidedly
- 4 quite
- 5 rather
- 6 somewhat
- 7 slightly

These modifiers may be used with adjectives to capture fairly evenly distributed categories. For example: Extremely (Sensitive, Valuable, Useful), Unusually (Sensitive, Valuable, Useful), and so on.

Flow Diagram for Ordered Category Analysis

If there are just two categories using ones and zeros, then **ORDER** Analysis or **GUTTMAN** Scaling may be appropriate. In handling unidimensional scales, **TSCALE** should be used. For most data, correlations and distances are the initial first calculations. These can be followed by a variety of methodologies. The authors prefer the SAS software system but SPSS, MiniTab or a variety of software may be used to compute correlations and do factor analysis.



Free Clustering

Steps in Free Clustering

In this task, the psychological objects (usually words, statements or concepts) are individually listed on slips of paper or cards. The participants are asked to put the objects they imagine or believe are similar into the same group (a single object may be a group). The respondents can constitute as many or as few groups as they feel are necessary. Free clustering is valuable because the underlying structure of the objects is not predetermined.

On the back of each card is written an identifying number (consecutive integers from 1 to k objects are used). When the groups have been formed, a new, smaller set of group numbers is assigned. Each different object in a particular group receives the same specific group number. For example, suppose five different judges are each presented the letters **a**, **b**, **c**, **d**, **e**, **f**, each one on a different card. The respondents group the letters together based on their estimates of letter similarity. Then the data might look as follows:

Group Numbers for Letters						
Judges	a	b	c	d	e	f
K	1	1	2	1	2	3
L	2	1	2	1	2	1
M	1	3	2	3	1	2
O	1	1	2	1	1	3
P	1	1	2	1	2	2

The integers in the data matrix represent group numbers. The percent overlap between each pair of letters can then be determined. **a** and **b**, for example, are in the same group three out of five times or .60. (See Measures of Proximity, p. 37). A matrix of similarities can be determined for these six lowercase English letters. This is accomplished by finding the number of times any two letters are found with the same group numbers and then dividing this sum by the number of subjects. A percent overlap matrix for the five judges in the example is provided in Table 2.2.

Table 2.2
Similarities Between Lowercase English Letters

	a	b	c	d	e	f
a						
b	.60					
c	.20	.00				
d	.60	1.00	.00			
e	.60	.20	.60	.20		
f	.00	.20	.40	.20	.20	

The program **PEROVER** (percent overlap), on the CD-ROM, calculates a similarity or distance matrix from the group membership data.

A quick perusal of the similarity matrix indicates that **b** and **d** are seen as most alike for these subjects since their similarity percentage is the highest (1.00).

Inter-Judge Differences

It is also possible to analyze the distances or dissimilarities between the judges. The differences are found by pairing all the *k* objects, for each judge, and recording a 1 if a pair is in the same group and a 0 otherwise. The judges' vectors are compared two at a time and the sum of the absolute value of the difference is determined.

The calculation of the differences for specific judges K and L based on their group membership numbers (K membership numbers are 112123 and L numbers are 212121) is shown below:

Judge	Letter Pairs														
	ab	ac	ad	ae	af	bc	bd	be	bf	cd	ce	cf	de	df	cf
K	1	0	1	0	0	0	1	0	0	0	1	0	0	0	0
L	0	1	0	1	0	0	1	0	1	0	1	0	0	1	0
d	1	1	1	1	0	0	0	0	1	0	0	0	0	1	0

$\Sigma|d| = 6$

The formulas for the mean and variance of the distribution of differences for *k* objects are:

$$\mu = k - 2 + 1/k \text{ and } \sigma^2 = k - 4 + 7/k - 6/k^2 + 2/k^3, \text{ where } k = \text{the number of objects.}$$

Dunn-Rankin & Wong (1980) wrote a computer program **JUDGED** to determine interjudge distances from the paired ratings of the *k* objects. This program is provided on the CD-ROM.

A single difference can be tested for its chance occurrence using a Z-test. Unfortunately, the normal distribution may not be appropriate because the clustering task implies fewer clusters than objects. The exact distributions have not been enumerated beyond *k* = 3

$\Sigma d $	*cp <i>k</i> = 3 Sample 20,000	cp <i>k</i> = 3 Enumeration
0	1.0000	1.00
1	0.7813	.789
2	0.4839	.493
3	0.0537	.049

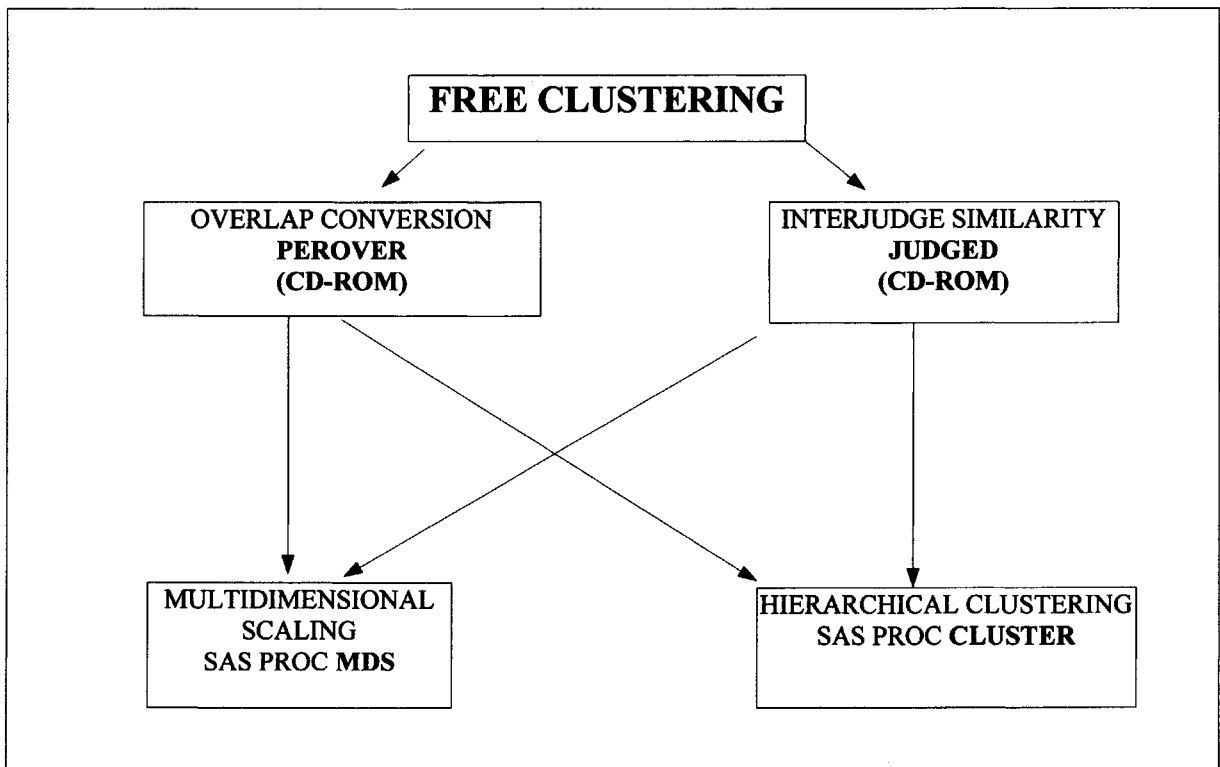
*cp = cumulative probability

Individualized Free Clustering

Individualized free clustering can be accomplished by listing the same objects in consecutive sets. The judges are asked to repeatedly mark or circle those objects that he or she would group together for a particular reason and to indicate, next to each grouping, the reason for the judged similarity. The respondents' reasons can be an important addition to the analysis. In this task, an object can appear in more than one group and overlapping clusters may result from the analysis of the similarity matrices. In this way, a percent overlap matrix can be constructed for each judge. Repeated free clustering in one individual is guided by the criteria consistent with the research question.

Flow Diagram for Free Clustering Analysis

A flow diagram for typical free clustering is provided below. After free clustering, a percent overlap matrix must be created (see Measures of Proximity, p. 37) the program **PEROVER** can analyze the raw group data and produce a similarity or distance matrix. The program **JUDGED** uses the same information and produces a matrix of distances between the judges. These programs are on the CD-ROM. The results can be input to SAS **CLUSTER** and **MDS**.



CD-ROM Example of Using PEROVER

letters.cfg (Configuration File)

```
5 6 1
letters.dat
letters.out
```

A configuration file lists the parameters and indicates the data input and output

letters.dat (Data Input File)

```
1 1 2 1 2 3
2 1 2 1 2 1
1 3 2 3 1 2
1 1 2 1 1 3
1 1 2 1 2 2
```

Data are group membership numbers for the five judges.

letters.out (Output File)

subjects = 5 variables = 6

raw data

```
1. 1. 2. 1. 2. 3.
2. 1. 2. 1. 2. 1.
1. 3. 2. 3. 1. 2.
1. 1. 2. 1. 1. 3.
1. 1. 2. 1. 2. 2.
```

The raw data

percent overlap matrix

```
1.000 0.600 0.200 0.600 0.600 0.000
0.600 1.000 0.000 1.000 0.200 0.200
0.200 0.000 1.000 0.000 0.600 0.400
0.600 1.000 0.000 1.000 0.200 0.200
0.600 0.200 0.600 0.200 1.000 0.200
0.000 0.200 0.400 0.200 0.200 1.000
```

Square similarity matrix

A comparison can be made with Table 1.2 (p. 23) the lower half similarity matrix constructed by hand. If a distance matrix was required the zero (0) in the configuration file would be changed to a one (1). In that case the matrix would look as follows:

Distance (difference) matrix

```
0.000 0.400 0.800 0.400 0.400 1.000
0.400 0.000 1.000 0.000 0.800 0.800
0.800 1.000 0.000 1.000 0.400 0.600
0.400 0.000 1.000 0.000 0.800 0.800
0.400 0.800 0.400 0.800 0.000 0.800
1.000 0.800 0.600 0.800 0.800 0.000
```

Square distance matrix

CD-ROM Example of Using JUDGED

letters.cfg (Configuration file)

5 6
letters.dat
letters.out

letters.dat (Input Data File)

1 1 2 1 2 3	K	} Judges
2 1 2 1 2 1	L	
1 3 2 3 1 2	M	
1 1 2 1 1 3	O	
1 1 2 1 2 2	P	

letters.out (Output file)

Number of subjects = 5 Number of variables = 6

1. 1. 2. 1. 2. 3.
 2. 1. 2. 1. 2. 1.
 1. 3. 2. 3. 1. 2.
 1. 1. 2. 1. 1. 3.
 1. 1. 2. 1. 2. 2.

Raw Data

Pop distance Mean = 4.17 variance = 3.01

Mean and Variance

Matrix of interjudge distances

0. 6. 5. 4. 2.
 6. 0. 5. 8. 8.
 5. 5. 0. 5. 5.
 4. 8. 5. 0. 6.
 2. 8. 5. 6. 0.

The square matrix output is a distance matrix and the first value, the distance between Judge L and Judge O or P is equal to 8, which is significant at the $P < .05$ level according to tabled sources such as Cumulative Probability Distributions of Interjudge Distances (Dunn-Rankin, 1983). One can also compare judges using a range test for outliers (Dixon & Massey, 1969).

Similarity Judgments

Paired Comparisons

In paired comparisons, the judges are presented the objects in all possible pairs and asked to make judgments of **similarity**. This may be in addition to the preference **choices** in each of the pairs.

Example: Five types of fishing lures are presented in all possible ways of choosing two objects out of k items or $[k(k - 1)/2] = 10$ ways. The judges provide values of **similarity** (S_{ij}) for each pair.

The judges also indicate their preference in each pair by checking the preferred item. In the example below, to the right of each pairing, is a number which reflects how similar Judge A thought each pair of lures happened to be. In this case 1 means not very similar and a 7 means very similar.

Lure Selection by Judge	Similarity
Rapala -*Lolo	3
*Rapala -Heddon	2
*Rapala -Windcheater	3
*Rapala -Kaku	2
*Lolo -Heddon	5
*Lolo- Windcheater	4
*Lolo -Kaku	4
*Heddon -Windcheater	2
Heddon -*Kaku	3
Windcheater -*Kaku	1

A similarity matrix can be established among the five lures for each judge. Judge A's responses are tabled below:

	Rapala	Lolo	Heddon	Wind Cheater	Kaku
Rapala					
Lolo	3				
Heddon	2	5			
Wind Cheater	3	4	2		
Kaku	2	4	3	1	

In the example the starred lures (*) were chosen as the best for catching fish in each pairing by Judge A. Using the asterisks, the votes are counted for each lure. A rank order for the lures results from the sum of the votes. Ties can occur due to circularity (see Circularity Triad Analysis, p.66). The votes by Judge A are as follows:

Ra	Lo	He	WC	Ka
3	4	1	0	2

Two programs on the CD-ROM, **INDMAT** and **AVEMAT**, are used to convert the basic data into matrix form for each subject or to average a group of subjects' responses.

Ranking Pairs

This task consists of ranking pairs of objects whose subsequent scores can be used in multidimensional scaling. For example, if the objects are **man**, **woman**, **boy**, and **girl**, six pairs are created and the pairs ranked on judgments of similarity (1 most similar, 6 least similar):

Pairs	Similarity Rank
man—woman	4
boy—girl	3
boy—man	2
boy—woman	5
girl—man	6
girl—woman	1

These ranks may be used, in a matrix of the objects, as direct estimates of distance. For example:

	Man	Woman	Boy	Girl
Man				
Woman	4			
Boy	2	5		
Girl	6	1	3	

Rating Similarity Between Pairs

Estimating similarities between pairs of objects can take the form on the bottom left and result in a matrix of similarities shown below right. One respondent's data is recorded. This simple matrix or a set of matrices can be explored multidimensionally. Several matrices can be averaged and analyzed.

	Degree of Meaning Similarity						
	Least Similar						Most Similar
Word Pairs	0	1	2	3	4	5	6
rose-rows			X				
rose-date				X			
rose-ate		X					
rows-date	X						
rows-ate	X						
date-ate							X

=

Word Similarity Matrix				
	rose	rows	date	ate
rose				
rows	2			
date	3	0		
ate	1	0	6	

Clustering Then Pairing

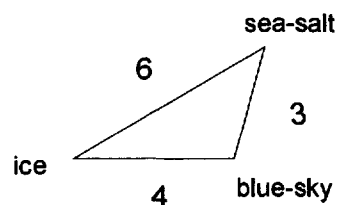
Once a number of objects have been clustered into groups, the groups may be paired and measures of similarity assigned to the paired groups. If, for example, one initially clusters the following words: ice, blue, sea, sky, and salt as follows:

Cluster	
1	Sea—salt
2	ice
3	Blue—sky

The clusters may be paired and rated in similarity. For example:

Pairs	Similarity	Distance
(sea—salt) — ice	4	6
(sea—salt) — (blue—sky)	6	3
ice — (blue—sky)	3	4

Next the similarities are converted to distances by subtracting from a constant larger than the maximum rating, say 10. Then a picture of the relationship between the three clusters can be drawn



This example can be expanded to include many items or objects. Clustering prior to pairing is a way of reducing the number of objects so that using paired comparisons becomes a manageable technique. Usually the centroid or average object of a cluster is chosen to represent the group in the pairing. In Euclidean space, the centroid is found by averaging the coordinates of the objects. This requires that a set of axes be established in which the set of objects is embedded.

Triadic Comparisons

In making triadic comparisons, the objects are formed in all possible triples. In this situation, the participant is asked to judge which two objects of each triple are most alike and which two are least alike. A set of scores can then be formed from the number of times that any object is judged more like one of a pair. Example: Four objects—**chuck**, **chock**, **check**, **chunk**—formed in all possible $[k(k-1)(k-2)]/6 = 4$ triples.

Four Triples

1	chuck	chock	check
2	chuck	chock	chunk
3	chuck	check	chunk
4	chock	check	chunk

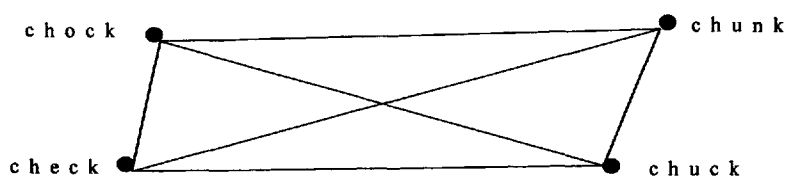
The paired triples can be handled like other paired data. Such data are useful in looking at the dimensions of **relationships**. Suppose the judgments from a **single** subject are as follows:

Pairs of Triples	Most Alike	Least Alike	Remaining
1	Chock—check	Chuck—chock	Chuck—check
2	Chuck—chunk	Chunk—chock	Chuck—chock
3	Chuck—chunk	Chunk—check	Chuck—check
4	Chock—check	Check—chunk	Chock—chunk
Score	2	0	1

A matrix (array of scores) for each subject can be formed in which choices are weighted, giving 2 points for pairs most alike, 0 for pairs least alike, and 1 for the remaining pairs. The summed scores represent measures of similarity. By subtracting the similarity score from a constant (in this case 5, one more than the maximum similarity), distances can be formed.

Similarity					Distance (5 - Similarity)				
	chuck	chock	check	chunk		chuck	chock	check	chunk
chuck					chuck				
chock	1				chock	4			
check	2	4			check	3	1		
chunk	4	1	0		chunk	1	4	5	

In this case, the objects may be approximately arranged in two dimensions as follows:



The scores for a number of respondents may be summed or averaged and then analyzed in a variety of ways. The data may also be handled as proportions (Torgerson, 1958).

Ratio Estimation

Another method of estimating similarities is **ratio** estimation (Ekman, 1963). In this method, the judge is asked to estimate what percent of stimulus "A" is contained in "B" and conversely what percent of "B" is contained in "A." Suppose, for example, the letters **o**, **b**, and **l** are used as stimuli. Judges are then asked, "Graphically, how much of **o** is contained in **b**, how much of **l** is in **b**," and so on. The ratios are as follows: o in b, b in o, o in l, l in o, b in l, l in b. The results for a single judge are given in the following matrix:

	First Judge		
	o	b	l
o		.40	.00
b	.90		.50
l	.05	.50	

The matrix is read, "How much of the column object is contained in the row object?" Thus the judge estimated 90% of **o** was contained in **b** but only 40% of **b** was contained in **o**. This array of data is therefore **asymmetrical** because the column by row entries are not equal to the row by column values. Such array data are generally averaged (using the mean or median) over a number of subjects. The resulting mean ratios are then converted to scalar products (see page 45) and factor analyzed. There are sophisticated methods for handling asymmetrical sets of data (Harshman, 1978).

Conditional Ranking

In conditional ranking one of the objects or stimuli is taken as a standard (S). Judges are asked to choose which of the (k - 1) remaining objects is most similar to the standard, then which of the (k - 2) objects is most similar, and so on until all of the objects have been ranked. Next, another object is randomly chosen as a standard and the process repeated. One can create a conditional rank order square distance matrix in this way. Each element of the matrix represents the rank similarity of object j to standard i. This may create an asymmetric (upper and lower triangles are not equal) set of data. Zeros on the diagonal assume a stimulus is most like itself. In the example below a 1 indicates very similar and 3 dissimilar. One person's responses are represented.

		apple	peach	orange	pear	
(S) apple-	pear(1), peach(2), orange(3)	apple	0	2	3	1
(S) peach-	pear(1), apple(2), orange(3)	peach	2	0	3	1
(S) orange-	peach(1), pear(2), apple(3)	orange	3	1	0	2
(S) pear-	apple(1), peach(2), orange(3)	pear	1	2	3	0

The upper and lower triangles of the matrix may be analyzed separately or the corresponding entries (i, j) and (j, i) can be added and averaged if the values are similar.

Same—Different

In perception experiments paired objects are briefly exposed visually or aurally and the participants are asked to determine whether the two objects are the same or different. The proportion of errors in judgment acts as a measure of similarity between the paired objects. For example **trial** and **trail** are presented together for 50 milliseconds and the judge is asked whether the words were the same or different. Or the two words **clothe** and **clove** might be spoken on tape and the respondent asked if the words were the same or different.

Latency

The length of time it takes to respond is sometimes used as a measure of similarity. In perceptual research, this usually requires some sort of shutter and a timing device. Computer programs now display data like a Tachistoscope. Two objects (letters, words, pictures, etc.) are hidden. When the stimuli are exposed a clock starts. The subject is asked to indicate whether two stimuli are the same or different; similar to a target or not by pressing one of two keys. Once a key is pressed the clock stops. If the average latency, for example, for judging whether **r** and **n** are the same is 400 milliseconds and the latency between **l** and **n** is 250 milliseconds one may assume that **n** is more similar to **r** than it is to **l**. If all pairs are used, a similarity matrix can be established.

Ranking Versus Rating Pairs

Because the most widely used methods for obtaining **direct** judgments of similarity are ranking and rating, it is worthwhile to investigate whether or not the separate scaling of ranking and rating responses to the same pairs of psychological objects achieves the same results. The rating task usually involves assigning estimates of similarity between each pair of objects on a scale of some stated range, whereas ranking calls for the ordering the pairs of objects in terms of increasing or decreasing similarity. Although ranking maximizes the transitivity (order) of the objects, it is the more demanding of the two tasks (takes more time). The difficulty it presents increases dramatically with an increase in the number of psychological objects. Rating, on the other hand, seems to be less demanding, requiring a shorter time for completion.

In order to determine whether rating can substitute adequately for ranking pairs of objects, judgments of similarity using both rating and ranking methods were compared. One such study was done using all pairs of seven campus places (Villanueva & Dunn-Rankin, 1973). The seven campus "places" were: **classroom**, **dormitory**, **library**, **cafeteria**, **gymnasium**, **theater**, and **laboratory**. The sequence of pairs, as they appeared on a rating sheet and in the pile of pairs followed the optimum presentation order recommended by Ross (1934) and was the same for both tasks and for all subjects.

For the rating task, each subject was instructed to assign any number from zero to 100 to each pair according to how similar the two places appeared. The most similar pair would have the highest rating, whereas the least similar would have the lowest rating. The 100-point rating scale was chosen because of its familiarity and relatively wide range.

Rating: The seven campus units yield 21 pairs of places. In the example below, each pair has been rated in similarity by one subject.

class-dorm	5	dorm-library	6	library-cafe	7	cafe-gym	2	gym-theater	6	Theater-Lab	4
class-library	24	dorm-café	4	library-gym	2	cafe-theater	8	gym-Lab	9		
class-cafe	2	dorm-gym	8	library-theater	19	cafe-Lab	3				
class-gym	34	dorm-theater	9	library-lab	6						
class-theater	27	dorm-lab	4								
class-lab	47										

Ranking: For the ranking task, there were 210 pairings. Each judge was instructed to sort the pairs initially into two piles consisting of (1) similar and (2) dissimilar pairs of pairs of objects. This initial sorting procedure had previously been found effective in simplifying the ranking task (Green & Carmone, 1970). Each respondent was instructed to order the pairs in both piles by circling the most similar of paired "campus places" on each card, that is, the votes for each pair of pairs was determined. The similarity configurations derived from both methods were then matched and compared. The authors concluded, with sophisticated judges and familiar objects, that both methods produce similar configurations under multidimensional scaling. Rating pairs can be substituted for ranking the pairs of pairs. Under similar conditions, the use of the rating method will economize the time and effort needed for tasks requiring judgment of similarity.

Analysis of Similarities

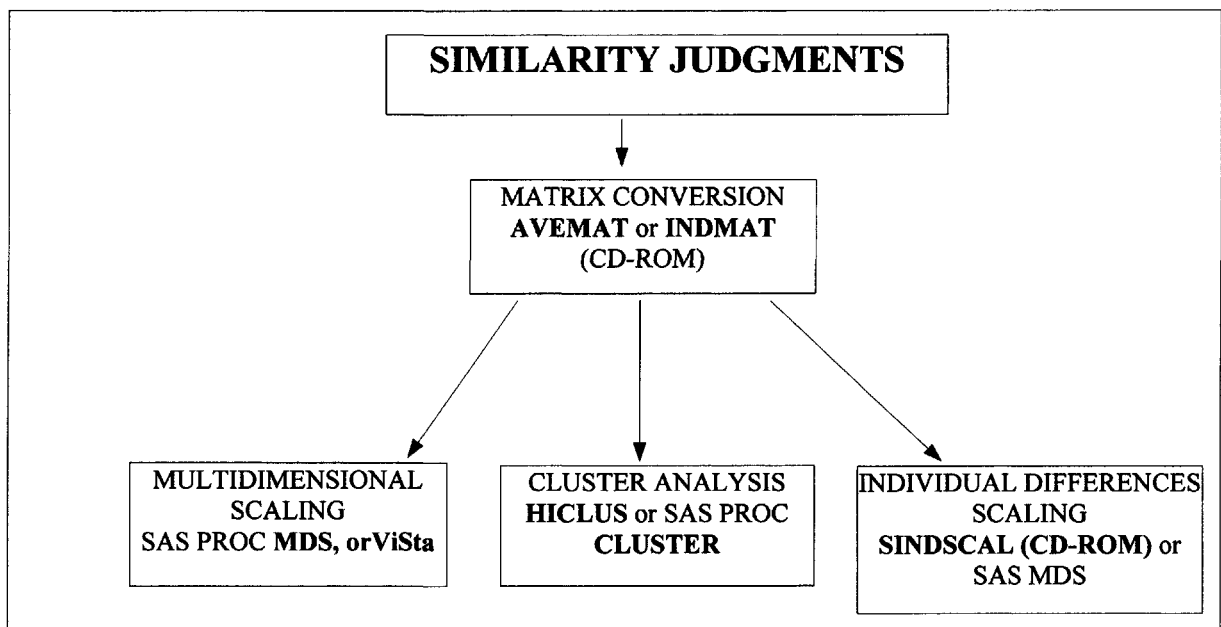
Similarity judgments can be handled by **CLUSTER** analysis or by multidimensional scaling (**MDS**). First, however, the data, which usually results from a number of pairings, must be coded with a key line indicating the order of the pairs. The data are a row (subjects) by column (similarity between each pair) matrix. This paired data is then put into either a square or half diagonal matrix.

For general analyses, all the examinees' data is averaged and usually converted into distances by subtracting them from the largest similarity value, that is usually a number found on the diagonal of the similarity matrix (the similarity of the object with itself). To do this, auxilliary programs on the CD-ROM such as **AVEMAT** or **INDMAT** are utilized. If the researcher is interested in individual differences then individual differences scaling using **SINDSCAL** on the CD-ROM or **SAS MDS**. These programs utilize an individual matrix for each subject.

When using direct judgments of similarity, the objects can be paired in a variety of ways (1 with 2, 8 with 5, etc.). The pairing code is provided by the researcher as part of the format of the auxiliary programs.

Flow Diagram for Similarity Judgments

Similarity judgments must initially be put into matrix form and then analyzed.



CD-ROM Example of AVEMAT

avemat.cfg (Configuration File)

```

5 4 7
avemat.dat
avemat.out
  
```

of Judges, # of objects, Maximum Similarity

avemat.dat (Input Data File)

```

01 02 01 03 01 04
02 03 02 04 03 04
5 2 4 2 2 5      DS
6 2 4 3 2 7      HA
4 2 4 4 2 4      DB
6 4 5 3 3 6      BB
5 3 6 3 2 5      AF
  
```

key code

data, ID is ignored

avemat.out (Output File)

Variables = 4 Subjects = 5

```

7.000 5.200 2.600 4.600
5.200 7.000 3.000 2.200
2.600 3.000 7.000 5.400
4.600 2.200 5.400 7.000
  
```

Square matrix of Similarities

CD-ROM Example of INDMAT**indmat.cfg** (Configuration File for INDMAT)

```

5 4 7
indmat.dat
indmat.out

```

indmat.dat (Input File for INDMAT)

```

01 02 01 03 01 04 02 03 02 04 03 04
5 2 4 2 2 5
6 2 4 3 2 7
4 2 4 4 2 4
6 4 5 3 3 6
5 3 6 3 2 5

```

indmat.out (Output File for INDMAT)

Variables = 4 subjects = 5

```

7.000 5.000 2.000 4.000
5.000 7.000 2.000 2.000
2.000 2.000 7.000 5.000
4.000 2.000 5.000 7.000

```

```

7.000 6.000 2.000 4.000
6.000 7.000 3.000 2.000
2.000 3.000 7.000 7.000
4.000 2.000 7.000 7.000

```

```

7.000 4.000 2.000 4.000
4.000 7.000 4.000 2.000
2.000 4.000 7.000 4.000
4.000 2.000 4.000 7.000

```

```

7.000 6.000 4.000 5.000
6.000 7.000 3.000 3.000
4.000 3.000 7.000 6.000
5.000 3.000 6.000 7.000

```

```

7.000 5.000 3.000 6.000
5.000 7.000 3.000 2.000
3.000 3.000 7.000 5.000
6.000 2.000 5.000 7.000

```

In this output there is one matrix of similarities for each subject

The averages from AVEMAT are used as input to HICLUS and MDS by editing out the unwanted lines and adding the SAS input code. The matrices from INDMAT may be used in Simplified Individual Differences Scaling, **SINDSCAL**.

3

MEASURES OF PROXIMITY

Proximities are numbers that indicate how close or how far apart objects appear. If two objects are very similar and this similarity is represented by a large number then such data are called similarity data or similarities. However, if two objects are very similar and this similarity is represented by a relatively small number, such data are inferred to be dissimilarities or distances. Most data in scaling methodology needs to be represented as distances so they can be mapped. If data are similarities they are usually converted to distances before any detailed analysis.

A number of proximity measures are available that relate to the tasks that have been previously presented. These can be categorized as measures of: (1) correlation; (2) distance; and (3) association. In order to analyze a set of objects through clustering or multidimensional scaling, some measure of similarity or dissimilarity between all the pairs of the objects is usually needed.

Correlations

Pearson's Correlation (r)

The Pearson Product-Moment Correlation (r) is a common measure of similarity. Pearson's r is defined as the average cross product of standardized scores or:

$$r = \frac{\sum Z_x Z_y}{N} \quad \text{or alternately} \quad r = \frac{\sum Z_x Z_y}{(N - 1)}$$

Where: Z_X and Z_Y are standardized scores and N is the number of scores and

$$Z = (X - \overline{X})/N .$$

This definition is important because it illustrates that all the scores that are handled by Pearson's r are standardized. This standardization assumes that (1) the data are samples from normal population with (2) similar variances and (3) X and Y are linearly related.

The raw score formula for r is as follows:

$$r = \frac{\sum XY - \sum X \sum Y / N}{\sqrt{[\sum X^2 - (\sum X)^2 / N][\sum Y^2 - (\sum Y)^2 / N]}}$$

The formula can be applied to an N by K (subjects by objects) set of raw data to generate a square K by K (object by object) symmetric matrix of correlations (**R**). (A symmetric matrix is a square array of numbers whose entries above the diagonal equal those below the diagonal.) Because information in a symmetric matrix is redundant, sometimes only the upper or lower triangle of entries is used. The results are graphically illustrated as follows:

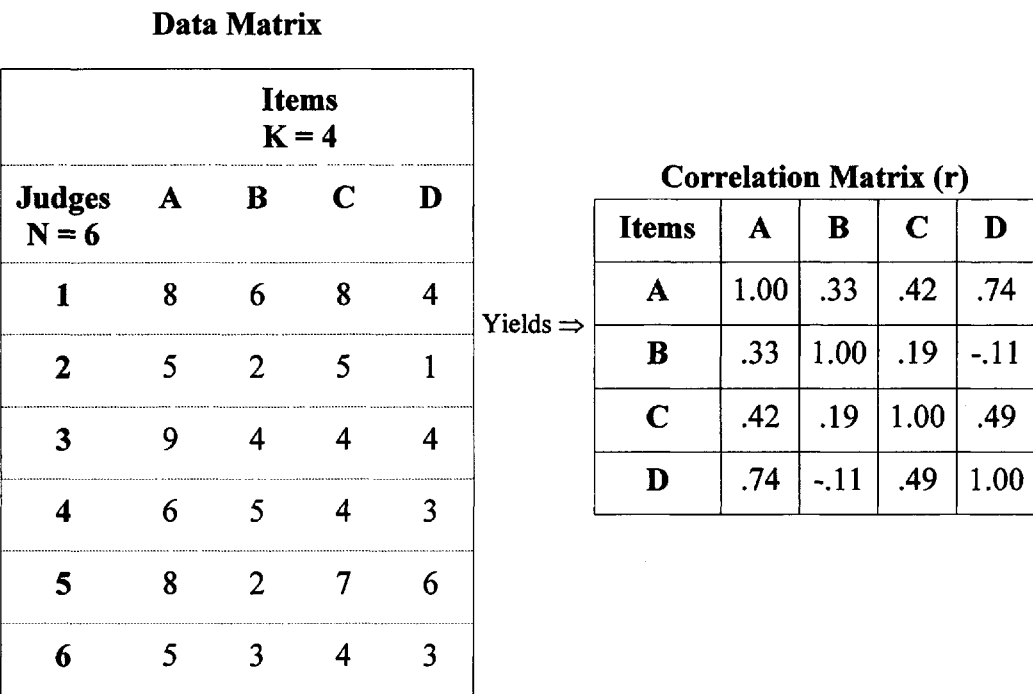


FIG. 3.1. An illustration of a raw data matrix producing a similarity matrix of correlations (r) between items.

SAS Example of Calculating Correlations

There are number of software packages that calculate correlations. The SAS (1999) system is representative. It is applied to the data of Fig. 3.1. The results provide the matrix of correlations as well as levels of significance.

SAS Input

```

data reexample;
input subjects I1-I4;
datalines;
01 8 6 8 4
02 5 2 5 1
03 9 4 4 4
04 6 5 4 3
05 8 2 7 6
06 5 3 4 3
;
proc corr;
run;

```

SAS Output

The CORR Procedure

4 Variables: I1 I2 I3 I4

Simple Statistics

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
I1	6	6.83333	1.72240	41.00000	5.00000	9.00000
I2	6	3.66667	1.63299	22.00000	2.00000	6.00000
I3	6	5.33333	1.75119	32.00000	4.00000	8.00000
I4	6	3.50000	1.64317	21.00000	1.00000	6.00000

Pearson Correlation Coefficients, N = 6
 Prob > |r| under H0: Rho=0

	I1	I2	I3	I4
I1	1.00000	0.33183 0.5205	0.41995 0.4071	0.74200 0.0913
I2	0.33183 0.5205	1.00000	0.18650 0.7235	0.07454 0.8884
I3	0.41995 0.4071	0.18650 0.7235	1.00000	0.48653 0.3278
I4	0.74200 0.0913	0.07454 0.8884	0.48653 0.3278	1.00000

Note that the value under each correlation is the probability of that value occurring by chance (if the null hypothesis of 0 r is true through random sampling in a population in which X and Y have no correlation)

Significance of r

The statistical significance of the Pearson correlation, its chance probability, can be determined by the following formula:

$$F = r^2 / (1 - r^2) / (N - 2)$$

Where r^2 is the square of the correlation and N is the number of pairs of objects. You can determine the alpha probability values for r in any "Critical Values for F" Table. Look under one (1) degree of freedom (df) for the numerator and N - 2 (df) for the denominator. Statistical significance means the absolute value of a correlation is so large that it wouldn't occur very often by chance. P = .05, for example, means that a specific correlation or one larger would occur only five times in 100 by chance.

By substituting the value of 4 for **F** (a value close to the 95% level) and solving for **r** in the formula above, a quick approximation of significant correlation at the .05 probability level can be obtained by using:

$$r_{.05} \cong 2 / \sqrt{N + 2}$$

If **N** = 14 the approximate correlation needed is .50, for example.

Squaring the Correlation Coefficient

When dealing with correlations it is wise to remember that a correlation can be statistically significant, that is, occur rarely by chance, but be of little predictive value. Squaring the By correlation coefficient (**r**²) tells us how much variance in one variable can be accounted for by the other. If a correlation between vocabulary and reading is **r_{vr}** = .50 then **r_{vr}**² = .25 and we can say that 25% of the variance in the scores in reading can be predicted by knowing scores on vocabulary. **r**² is sometimes called the coefficient of determination and 1-**r**² the coefficient of non-determination. Sometimes (1 - **R**²)^½ is used as a measure of predictability known as the “badness of fit”

Pearson’s **r** can be applied to binary data and to ranked data. When this happens the correlation has historically taken on different names. The Point Biserial Correlation (**r_{pt bis}**) measures the association between two variables one of which is continuous and the other a dichotomy. The tetrachoric correlation (**r_{tet}**) measures the association between two dichotomies which have been converted from continuous distributions and **Rho** is the correlation between ranks.

Pearson’s r Relationships	
Correlations Based on Pearson’s r	Measures based on r ²
r_{tet} Tetrachoric (Two Dichotomies)	r ² (Accountable Variance or Coefficient of Determination)
r_{pt bis} Point Biserial (One continuous and one dichotomy)	1 - r ² (Coefficient of Non Determination)
Spearman Rank Rho (Data are ranks)	(1 - r ²) ^½ (Badness of Fit)

Kendall's Tau Correlation

Kendall's (1952) **tau** has special relevance for measuring similarity because it is a rank order correlation coefficient with fewer assumptions than Pearson's **r**. It can be applied to ordered category scales and it forms the basis for other widely used measures of association, such as Goodman-Kruskal's gamma (**g**). In **r_{tau}**, all possible pairs of scores are compared for each variable separately. When + 1 is assigned to concordant pairs and -1 to discordant pairs, the **tau** coefficient can be calculated by finding the sum of the products of the concordant or discordant scores in the two sets of pairings and dividing by the number of possible pairs [N (N-1)/2]. (A pair is concordant if the numbers associated with each pair are in ascending rank order, otherwise the pair is discordant.)

Suppose the following rank data are provided for six students.

Rank Scores		
Students	Achievement	Motivation
1	2	3
2	1	2
3	3	4
4	4	6
5	6	5
6	5	1

Comparing students 1 and 2, in the table above, their Achievement ranks are not in descending order (2 is before 1). The same is also true for their Motivation ranks (3 over 2). In both cases a -1 is recorded. The product of these two cases, however, is concordant or $[(-1)(-1) = +1]$. The six students can be paired in 15 ways and the order of their scores shown as follows:

Students	1-2	1-3	1-4	1-5	1-6	2-3	2-4	2-5	2-6	3-4	3-5	3-6	4-5	4-6	5-6
Achievement	-1	1	1	1	1	1	1	1	1	1	1	1	1	1	-1
Motivation	-1	1	1	1	-1	1	1	1	-1	1	1	-1	-1	-1	-1
Product	1	1	1	1	-1	1	1	1	-1	1	1	-1	-1	-1	1

10 concordant ($N_c = 10$) and 5 discordant ($N_d = 5$) sums are recorded and S = the sum of the positive and negative 1's or $S = 10 - 5 = 5$.

$$r_{\text{tau}} = S/[n(n-1)/2], \text{ n is the number of subjects.}$$

$$= (5/15)$$

$$= 0.33$$

SAS PROC CORR KENDALL calculates **tau-b** correlations and provides tests of significance based on the normal distribution of r_{tau} , where

$$Z = r_{\text{tau}} \sqrt{9N(N-1)/2(2N+5)}$$

Tau-b correlations are corrected for ties by dividing S by the geometric mean of the number of pairs **not** tied in each of the two sets of scores. If $X = 2, 3, 2, 2$ and $Y = 2, 3, 3, 2$ the six ordered pairs are scored as follows:

X	+1	0	0	-1	-1	0
Y	+1	+1	0	0	-1	-1
XY	+1	0	0	0	+1	0

a = number of pairs not tied = 3
 b = number of pairs not tied = 4 $ab = 15$
 $S = N_c - N_d = 2 - 0 = 2$

$$\text{Tau-b} = 2/(ab)^{1/2}$$

$$= 2/(12)^{1/2} = 0.578$$

$$r_{\text{tau}} \text{ (without correction)} = S/\text{number of pairs}$$

$$= 2/6 = 0.33$$

r_{tau} will usually provide smaller sized coefficients than Pearson's r even though the statistical significance may be the same.

Gamma Correlation

Gamma (g) is simply the number of concordant products minus the number of discordant products divided by the sum of the concordant and discordant products or

$$g = (N_C - N_D) / (N_C + N_D)$$

If there are no ties in the data, gamma will equal tau. If ties are present gamma will be greater than tau. Usually there are a great many ties in data that are to be ranked. Gamma will, therefore, provide larger indices of proximity but will use less of the data.

Distances

Sometimes respondents are intercorrelated using objects or variables as profile information. Care must be taken in interpreting profile results because two judges can be highly correlated (because their profiles have the same pattern) yet differ widely in the level of their scores. This is true, for example, with subjects 01 and 02 in Fig. 3.1.

Judge 1	8 6 8 4
Judge 2	<u>5 2 5 1</u>
Difference	3 4 3 3

In this case the difference between the two profiles is relatively large yet they correlate highly $r = .97$. Should such differences in profiles be important, other measures must be obtained, such as the sum of the absolute differences

$$\sum |X_i - X_j|$$

or some measure of Euclidean distance.

Measures of distance are dissimilarity measures of proximity. They can, for example, solve the problem of similar profile patterns that vary in magnitude. The distance measure between two points in a plane is

$$d_{ij} = [(X_i - X_j)^2 + (Y_i - Y_j)^2]^{1/2}.$$

Standardized Distances

Distance calculations are often applied to a profile of standardized scores. After the raw data

$Z_i = (X_i - \bar{X})/S_x$ where S_x is the standard deviation

Mahalanobis d^2

The Mahalanobis d^2 statistic standardizes the variables before calculating Euclidean distance by dividing each dimensionally squared difference by the variance of that variable. If i and j are variables then:

$$d_{ij}^2 = (X_i - X_j)^2/S_x^2 + (Y_i - Y_j)^2/S_y^2$$

Minkowski Metric

A general formula for distance, called the Minkowski metric, can be written as follows:

$$d_{ij(p)} = \sum [|X_{it} - X_{jt}|^p]^{1/p} \text{ Where } t \text{ runs from } 1 \text{ to } r \text{ dimensions.}$$

When p (the power) = 2 the formula is equal to Euclidean distance. In this case, $1/p$ becomes $1/2$ or another symbol for the square root. If $p = 1$, the formula reduces to:

$$d_{ij} = \sum |X_i - X_j|$$

commonly known as the city block metric. This is so named because to go from one corner of a block in the city to the diagonally opposite corner you can't go through the buildings and must go around the block.

Using the data from Fig. 3.1, Euclidean distances were calculated using the Minkowski Metric with $p = 2$. Note that the distances between the objects are different from the correlations. Where objects A and D are highly correlated (.74); they are relatively far apart (8.6) when distances are calculated.

Correlation Matrix (r)				
Items	A	B	C	D
A	1.00			
B	.33	1.00		
C	.42	.19	1.00	
D	.74	-.11	.49	1.00

Euclidean Distance Matrix				
Items	A	B	C	D
A	0.0			
B	8.8	0.0		
C	5.6	4.5	0.0	
D	8.6	5.0	5.9	0.0

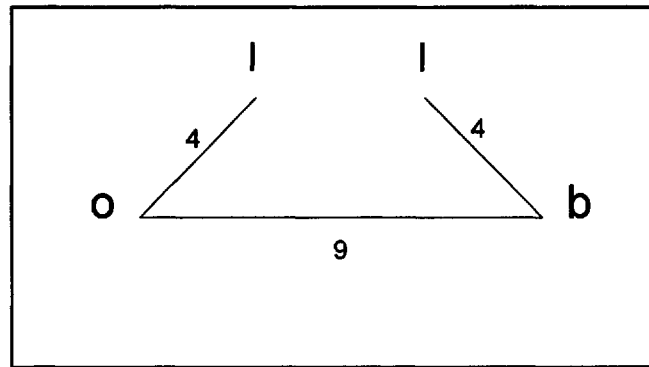
Triangle Inequality

When using distances, most methods of analysis must satisfy the following properties:

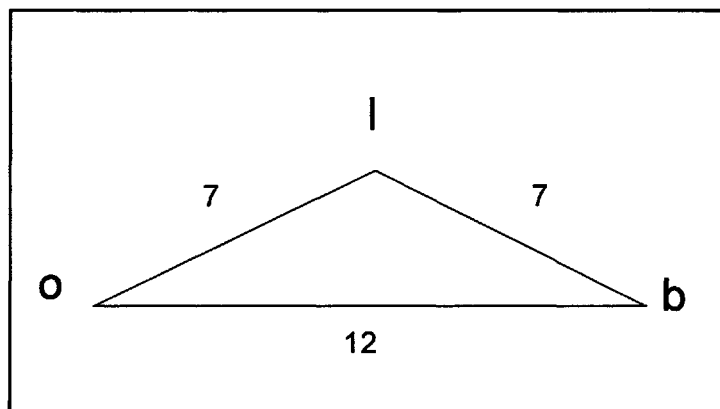
1. The distance between any two objects is symmetrical (i.e., $d_{xy} = d_{yx}$).
2. The distance between x and y is zero only if $x = y$.
3. If x , y , and z are three objects, the distances between the three objects must form a triangle (i.e., $d_{xz} < d_{xy} + d_{yz}$).

This last condition is called the triangle inequality. Sometimes when this condition is not satisfied a constant may be added to each value to fulfill the third condition. It is known, for example, that the psychological distance between the letters **o** and **l** is considered to be much larger than the sum of the psychological distances of **o** and **b**, and **b** and **l**.

Schematically:



Thus, in raw form, the distances do not form a triangle. By adding 3 to each distance the inequality can be solved as follows:



Usually the minimum constant required to satisfy the triangle inequality for all triples in an entire matrix is added to each distance. Note: adding a constant is suitable with data measured on the interval scale but not appropriate with data measured on a ratio scale with a meaningful zero.

Scalar Products

Distances and correlations can be functionally related by using scalar products. In order to understand this relationship, it is useful to review the law of cosines. Several multidimensional methods calculate the scalar product as an initial step in their analysis.

In a right triangle the cosine of the angle equals the length of the adjacent side (x) divided by the length of the hypotenuse (c).

The relationship between points B and C can be inferred by calculating the scalar product by just knowing the distances a, b, and c.

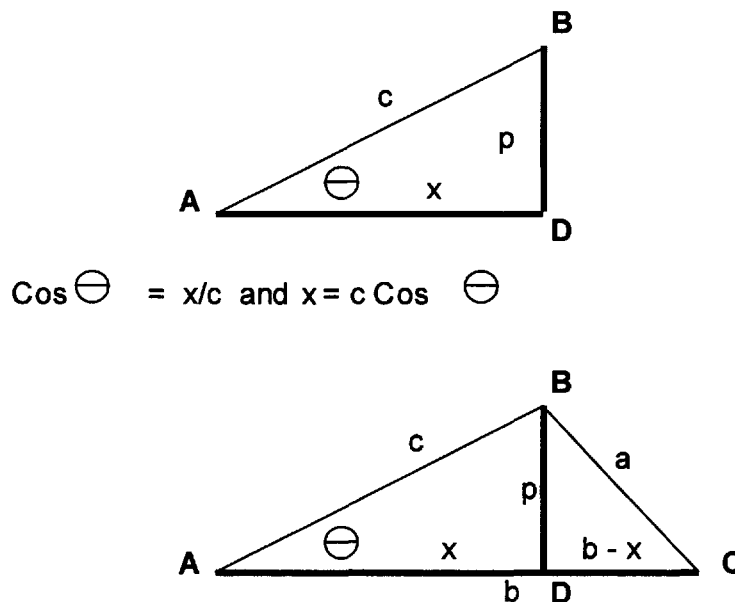


FIG. 3.2. Triangle connecting points A, B, and C.

- | | | |
|----|--|--|
| 1. | $c^2 = x^2 + p^2$ | In $\triangle ABD$ |
| 2. | $p^2 = a^2 - (b - x)^2$ | In $\triangle BCD$ |
| 3. | $p^2 = c^2 - x^2$ | |
| 4. | $a^2 - (b - x)^2 = c^2 - x^2$ | Substituting for p^2 in (2) |
| 5. | $a^2 = c^2 - x^2 + (b - x)^2$ | Squaring and Summing |
| 6. | $a^2 = c^2 + b^2 - 2bx$ | Rearranging terms |
| 7. | $a^2 = c^2 + b^2 - 2bc \cos \Theta$ | From Fig. 3.2 substituting $c \cos \Theta$ for x |
| 8. | $bc \cos \Theta = (c^2 + b^2 - a^2) / 2$ | |

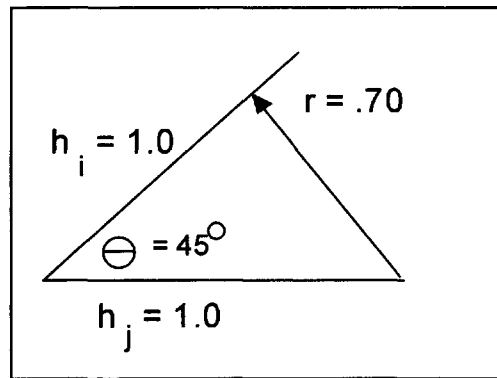
The distance (a) between two points (B and C) can be obtained by knowledge of the length of two vectors (that terminate at B and C) that have the same origin (A) and by knowing the size of the angle included between the two vectors.

The left-hand side of Equation 8 is called the scalar product and is often taken as a substitute for the correlation between the two points represented by the vectors. The correlation is sometimes written

$$r_{ij} = h_i h_j \cos \Theta_{ij}$$

when the data have been standardized (h = length of vector).

If two vectors are of unit (1.0) length and the angle between them is acute, then the cosine of the included angle should directly reflect their correlation. If a perpendicular line is drawn from the end of unit vector h_j to h_i , its projection point represents the scalar product.



An evaluation of the scalar product and therefore the correlation can be made with only the knowledge of the distance between the objects. Suppose, for example, one knows only the distances between objects i , j , and k as follows:

$$\begin{aligned} d_{ij} &= 10 \\ d_{ik} &= 8 \\ d_{jk} &= 6 \end{aligned}$$

Suppose also that interest is in the scalar product between j and k (call this value b_{jk}).

From step 8 we know that

$$\begin{aligned} b_{jk} &= \frac{1}{2} (d_{ij}^2 + d_{ik}^2 - d_{jk}^2) \\ &= \frac{1}{2} (10^2 + 8^2 - 6^2) \\ &= 64 \end{aligned}$$

Several methods in multidimensional scaling convert similarities to distances to scalar products because the algebra associated with their solutions is amenable to this measure of proximity.

The methods may also calculate the scalar product based on the coordinates of the points. This scalar product

$$b_{jk} = \sum x_{jt}x_{kt}$$

is known as the dot product (dot represents multiplication), and x_j and x_k are the vector coordinates of objects j and k on dimension t . t runs from 2 to r dimensions. The sum of the cross products of the coordinates, b_{jk} , provides the scalar product.

Association

Direct Estimation of Proximity

Usually direct estimates of similarity are gathered in some form of pairwise comparisons. In addition to asking for a preference or choice, degrees of judged similarity are sought.

Example: Given four objects A, B, C, and D paired in all possible (6) ways. If two students make direct estimates of similarity between each pair of objects, their raw data might look as follows:

	A-B	A-C	A-D	B-C	B-D	C-D
Student 1	7	4	1	5	2	6
Student 2	5	5	2	6	3	7

For each subject a matrix of direct similarities results.

Subject 1					Subject 2				
Object	A	B	C	D	Object	A	B	C	D
A					A				
B	7				B	5			
C	4	5			C	5	6		
D	1	2	6		D	2	3	7	

These data may be analyzed together by averaging the scores, or analyzed separately for each subject or both.

Percent Overlap

In a free clustering task, where judges are asked to group similar objects together, different judges may each include the same objects in the sets of objects they place together. Should two objects be frequently placed together by most or all of the subjects, the assumption is that they are similar. An index of this similarity is the proportion of times any two objects are found in the same group. Suppose for example, five subjects group the words **love**, **hate**, **happy**, **bitter** and **sad** as follows:

Subjects			
1	[love, hate]	[sad, happy]	[bitter]
2	[love, happy]	[sad, bitter]	[hate]
3	[love, hate, happy, sad]		[bitter]
4	[love, hate, bitter]		[sad, happy]
5	[love, sad]	[hate, bitter]	[happy]

Because love and hate are put together in a group by three of the five subjects the similarity score is 3/5 or .60.

A similarity matrix of the proportions for the foregoing data is as follows:

	love	hate	sad	happy	bitter
love	1.00				
hate	0.60	1.00			
sad	0.40	0.20	1.00		
happy	0.40	0.20	0.60	1.00	
bitter	0.20	0.40	0.20	0.00	1.00

Minimum Percentage

The sum of the estimated time or percentage between two variables or categories is simply

$\sum \text{Min}(X_i, Y_j).$

For example, the estimated time spent for various activities over a 24-hour period is presented in the following table:

	Ideal	Real	Minimum
Teaching	3	3	3
Recreation	3	3	3
Writing	1	1	1
Research	3	1	1
Eating	1	1	1
Sleeping	8	6	6
TV	2	5	2
Social	2	1	1
Travel	1	1	1
Misc.	0	2	0
Total	24	24	17

% Overlap = 17/24
= .71

The value of .71 represents the extent to which the person is approaching an ideal life or a value of 1.00.

Interjudge Differences Following Free Clustering

Often, judges will **not** cluster objects in similar ways. The researcher is interested in whether some judges should be excluded or separated from the rest and the analysis performed over separate subsets of judges.

One way to determine the difference between two judges is to pair the objects in all possible ways and create a vector of ones and zeros for each judge by providing a score of one (1) when paired group numbers are the same and zero otherwise (the same group numbers are given to all objects placed in the same group or cluster). The difference between two judges is simply the sum of the absolute or squared differences between their two vectors of ones and zeros.

For example, the five words **love**, **hate**, **happy**, **sad**, and **bitter** are paired in all possible 10 ways. A vector of ones and zeros is created for the first two subjects indicating whether or not they placed the pair together in their free clustering. The sum of the absolute difference vector equals the interjudge difference.

	love	love	love	love	hate	hate	hate	happy	happy	sad
Judge 1	1	0	0	0	0	0	0	1	0	0
Judge 2	0	1	0	0	0	0	0	0	0	1
Difference	1	1						1		1

D₁₂ = 4

The interjudge difference D₁₂ is equal to 4.

A FORTRAN program, **JUDGE**, has been written which calculates the differences between all the judges following free clustering and provides the mean and variance of the difference distribution. It is provided on the CD ROM.

Gower's Similarity Measure

Gower (1971 has defined a general measure of similarity between two subjects or objects that can combine qualitative, binary, and quantitative data. The measure can be used independently on a set of scores of a single type or on a mixture of scores. This index is:

Similarity_{ij} S_{ij} = ∑ S_{ij(k)} / ∑ W_{ij(k)},

or the similarity measure between objects i and j over k variables divided by the weight associated with each i, j, and k. The weights (w_{ijk}) are usually 1 for each variable so that the ∑W_{ijk} for all variables will ordinarily be the number of different variables. If a comparison is not valid, or variable k is unknown or absent for one or both of the objects being compared, then the w_{ijk} is equal to 0. The weight depends on what type of data is being analyzed. S_{ijk} is computed as follows:

Binary:	$S_{ij} = 1$ if values of variable are same for both objects.
Qualitative:	$S_{ij} = 1$ if same quality is present in both objects.
Quantitative	$S_{ij} = 1 - x_{ik} - x_{jk} / R_k$ where x is the score and R is the range of the variable k .

If, for example, respondents are measured on IQ, reading score, sex, and ethnic origin as follows:

Subjects	Quantitative		Binary	Qualitative
	IQ	Reading	Sex*	Ethnic Origin
1	100	50	1	Caucasian
2	90	50	2	Japanese
3	110	45	1	Japanese
4	115	60	2	Filipino
5	130	54	1	Caucasian

The similarity between each pair of respondents can be calculated. The similarity between persons 1 and 2 is used as an example.

Measure	Variable	$1 - x_{ik} - x_{jk} / R_k$	Similarity $S_{12(k)}$	Weight $W_{1,2(k)}$	Ratio $S_{1,2(k)} / W_{1,2(k)}$
1. Quantitative	IQ	$1 - 100 - 90 /40$	$= .75$	1	.75
2. Quantitative	Reading	$1 - 50 - 50 /15$	$= 1.0$	1	1.0
3. Binary	Sex		$= 0$	1	0
4. Qualitative	Ethnic Origin		$= 0$	1	0
			1.75	4	1.75

Gower's Index is:

$$S_{12} = \sum S_{1,2(k)} / \sum W_{1,2(k)} = 1.75/4 = .44$$

Table 3.1 presents all the pairwise measures of similarity for these data.

Table 3.1
Set of Gower's S_{ij} Measures

S	1	2	3	4	5
1					
2	.44				
3	.60	.54			
4	.24	.43	.22		
5	.75	.07	.48	.31	

A number of computer programs are useful in establishing measures of proximity (similarity or distance). Most programs translate the data from raw form (subject by object data) into some form of proximity matrix (usually object by object). The most common measure of similarity used is Pearson's r correlation. Not so common is the calculation of Kendall's tau although it is found in SAS and SPSS (see p. 30). SAS has a sample library, a % **Distance Macro** which contains a number of measures of similarity and distance. These include GOWER (Gowers's similarity measure), EUCLID (Euclidean distance), L(p) (Minkowski distance where p is a positive value), and many others including the MINRATIO (for binary data). A listing is provided as a guide. Measures of association are also available. **PEROVER** (a program on the CD-ROM) converts free clustering group data into distances. This auxiliary program is provided with this text along with the program **JUDGED** to determine interjudge differences.

Kappa

Kappa is a measure of inter-rater agreement and ranges from 0 to 1.

$\text{Kappa} = (\text{observed concordance}) - (\text{concordance by chance}) / 1 - \text{concordance by chance}.$

Suppose two raters rate 20 objects into two categories A and B. These are the marginal sums. The raters agree that 8 of the objects belong in category A and 6 in category B. The data might look as shown on the right.

		Rater 2		
Rater 1		A	B	Sum
	A	8	4	12
	B	2	6	8
	Sum	10	10	20

Observed concordance (agreement) = diagonal sum divided by N
 $= (8 + 6) / 20$
 $= .70.$

Expected concordance (chance) = the sum of each row total multiplied by its corresponding column total / N . That is $(12)(10) / 20 = 6$ and $(8)(10) / 20 = 4$. These are the expected diagonal frequencies. They are also summed and divided by N or $(6 + 4) / 20 = .50$

$$\begin{aligned}
 \text{Kappa} &= (\text{Obs} - \text{Ex}) / 1.0 - \text{Ex} \\
 &= (.70 - .50) / (1.0 - .50) \\
 &= .20 / .50 \\
 &= .40
 \end{aligned}$$

SAS supports the calculation of Kappa and calculates its standard error. Use

```
Proc Freq;
Table Judge1 Judge2 / kappa;
Run;
```

A Distance Macro From SAS

SAS provides a distance macro to calculate similarity or dissimilarity data in a number of ways. (See SAS V8). Suppose four judges have scores on five variables and the researcher wishes to determine the similarity or dissimilarity between the judges. The experimenter chooses the City Block Metric, that is, the Minkowsky metric when $p = 1$. The SAS input data set data might look as follows:

```
data distance;
input judge $ i1-i5;
datalines;

J1 2 4 6 8 2
J2 3 6 2 1 5
J3 1 3 5 8 3
J4 5 7 4 3 7
;
run;
%inc 'C:\Program Files\SAS Institute\SAS\V8\STAT\SAMPLE\xmacro.sas'
%inc 'C:\Program Files\SAS Institute\SAS\V8\STAT\SAMPLE\stdize.sas';
%inc 'C:\Program Files\SAS Institute\SAS\V8\STAT\SAMPLE\distnew.sas'
%distance(out=distout, id=judge,
method=L(1), var=i1-i5);
Proc print data=distout;
Run;
```

Locates the distance macro

The method is $L(p)$, $p = 1$
Prints the matrix

The results of $L(1)$ which is the sum of the absolute differences for each Judge is:

Judge	j1	j2	j3	j4
j1				
j2	17			
j3	4	17		
j4	18	9	18	

Methods that can be substituted for the City Block are: Euclidean distance, correlation, squared correlation, Minkowski metric, similarity ratio, and many others.

PART II

UNIDIMENSIONAL METHODS

Unidimensional Scaling attempts to represent similarity judgments or preferences using a straight line. Carol Pang (1996), for example, ordered over 6000 nouns, taken from Webster's college dictionary. She did this on the basis of the one dimension of familiarity. For example, **mom** is more familiar than **bicycle** is more familiar than **chef** is more familiar than **frigate** is more familiar than **cetacean**, etc.

More Familiar **mom > bicycle > chef > frigate > cetacean** Less Familiar

The words were then grouped into 32 levels of familiarity. It is possible to locate an individual's level of familiarity by displaying just five words, one at a time, and asking the person how well he or she understands each particular word. Using a computer and an interactive program (Berg, 1995) this binary sort can be repeated 20 times in just a few minutes with a high degree of reliability.

In Part II, five major unidimensional methods are presented. Despite recent advances in multidimensional scaling, unidimensional methods have value because of their simplicity, versatility, and history. They are also less exploratory than multidimensional methods. Their underlying simple structure is usually well and narrowly defined. Unidimensional scales may have greater construct validity. They are also amenable to hand-calculated solutions. Multidimensional analysis often ends with several separate unidimensional scales which can be separately tested. One methodology can serve as a check on the other.

Each of the unidimensional methods presented offers something unique to scaling analysis. With **Rank Scaling** it is simplicity and tests of significance. In **Comparative Judgment** it is meeting normality assumptions about attitude. **Scalogram analysis** provides an "order" definition of scaling. In **Successive Categories**, profile data are handled instead of paired data. In Successive Category scaling the effect of varying interval widths is explored. **Summated Ratings** assumes equal intervals and relies extensively on test statistics.

This page intentionally left blank

4

RANK SCALING

Variance Stable Rank Sums

The variance stable rank method of scaling (Dunn-Rankin, 1965, Dunn-Rankin & King, 1969) is an adaptation of a two-way analysis of variance by **ranks**. In other words it is a nonparametric (distribution free) subject by treatment analysis in which the treatments are the psychological objects that are scaled. The basic assumption of the method is that the scale values are proportional to the sum of the ranks assigned by the judges to each of the objects. In this method, the maximum and minimum possible rank totals, for a given number of judges and objects, act as a convenient and interpretive frame of reference within which the objects are scaled. A linear transformation of these two extreme rank totals into 100 and zero, respectively, defines the limits of the scale.

The psychological objects can be ranked directly or the ranks can be determined from the votes given to the objects when they are arranged in all possible pairs and a choice made of the most preferred of each pair. Complete ordering can also be derived from partial ranking procedures, that is, Balanced Incomplete Block (BIB) designs (see p. 12).

A group of second-grade children were asked what they most preferred as a reward after a job was well done; an **A**, a **100**, a **Gold Star (GS)**, or the word **Excellent (Ex)**. For these children the objects were formed in all $[K(K - 1)/2]$ or six possible pairs. The marked objects in Figure 4.1 indicate the preferred choice in each pairing for the first child. The figure also shows the preference values for child 1. These values are found by summing the votes for each different object. In this case, three choices or votes were made for a 100, two votes for an A, 1 vote for the Gold Star and **no** preference for the word Excellent. The 3, 2, 1, and 0 rank values are the reverse of the rank order of the objects but are utilized in this position so that the value associated with the most preferred object has the largest magnitude.

A — ♦100, Ex — ♦A, GS — ♦100, ♦A — GS, ♦100 — Ex, ♦GS — Ex

Ex	GS	A	100
0	1	2	3

FIG. 4.1. Reward pairings. Marked object (♦) was preferred in each pairing.

Table 4.1 shows the rank values obtained for 24 students over the same objects. After obtaining the rank value totals (R_k), the Scale Values (SV) are obtained by dividing each vote total by the maximum vote possible (R_{tot}) and multiplying by 100. These values and a unidimensional graph are presented at the bottom of Table 4.1.

Table 4.1
Calculation of Scale Values (SV) from Sum of the Rank Values

Students	Min	Ex	GS	A	100	Max
1	0	1	0	2	3	3
2	0	0	1	3	2	3
3	0	0	1	2	3	3
*4	0	1	1	1	3	3
5	0	3	2	0	1	3
6	0	0	1	2	3	3
7	0	1	0	3	2	3
8	0	0	3	2	1	3
9	0	0	1	3	2	3
*10	0	2	0	2	2	3
11	0	1	0	2	3	3
12	0	1	0	3	2	3
13	0	0	1	3	2	3
14	0	3	2	0	1	3
15	0	1	0	2	3	3
16	0	0	1	3	2	3
17	0	3	0	2	1	3
18	0	0	1	2	3	3
19	0	0	3	1	1	3
20	0	0	3	2	2	3
21	0	2	0	2	2	3
22	0	0	1	3	2	3
23	0	0	1	2	3	3
24	0	0	1	2	3	3
Sums (R_k)	0	19	24	49	52	72
$SV=(100R_k/R_{tot})$	0	26.3	33.3	68.1	72.2	100

*Circular triads present. For student 4, Ex is chosen over GS; GS over A, and A over Ex. In such cases the procedure assigns the mean of the three rank values to each of the items involved. $R_{tot} = N(K-1)$.

Graph of Reward Preference Scale Values

R_{\min}	Ex				GS				A				100	R_{\max}
0	20	26	30	33	40	50	60	68	72	80	90	100		

Test of Significance

When tests of significance are incorporated into the scaling process, **variance stable** scales of paired comparison data can be constructed. The rank method has units that are equal in a variance stable sense. They are **variance stable** because a specific difference between rank sums has the same probability of occurrence wherever the rank totals (and the scaled scores) may be located. In Table 4.1 the .05 critical range is a value of **23**. Any difference between rewards that is equal to or greater than 23 is statistically significant. In scaling, we are usually interested in all the possible comparisons that can be made between the objects. Utilizing rank totals in the scaling process provides the opportunity to examine significant differences between pairs of psychological objects.

The nonparametric method of multiple comparisons (Wilcoxon & Wilcox, 1964) concentrates on detecting differences between treatments and in this case is analogous to the Tukey method of multiple comparisons (Hays, 1973). This method uses the range and declares that objects with significantly large **rank sum differences** will have significantly different means.

Dunn-Rankin and Wilcoxon, 1966 investigated the true distribution of the range of rank totals, that is, the distribution of the largest differences between the rank sums. The study determined the exact probability values necessary when using small numbers of judges and objects. Table D in Appendix B presents critical values of the rank sum differences where K (objects) and N (judges) are less than or equal to 15. The study was also able to verify that the normal approximation to this distribution is accurate when the number of judges or objects is greater than 15. Thus, the probability can be easily calculated for N or K > 15.

Calculation of the critical range for the sample of 24 subjects shown in Table 4.1 is illustrated below. The critical range is the product of the expected standard deviation, E(S) and a value from the range distribution, Q_a .

$$E(S) = \sqrt{N(K)(K+1)/12}$$

Where K = the number of objects and N = the number of judges.

If K = 4 and N = 24 then

$$\begin{aligned} E(S) &= \sqrt{24(4)(5)/12} \\ &= \mathbf{6.325} \end{aligned}$$

Q_a = W/S is the Studentized range for K treatments and infinite df. For N = 24, K = 4 and P = .05, the value, 3.633, is obtained from Table C in Appendix B.

$$\begin{aligned}
 \text{Critical Range} &= E(S) Q_{.05} \\
 &= (6.325)(3.633) \\
 &= 22.978 \\
 &= 23
 \end{aligned}$$

An illustration of the calculation of the significant range for 24 judges and four objects where the .05 probability level is chosen is provided. The values of the Studentized range ($Q_\alpha = W/S$) have been extensively tabled by Harter (1959) or may be found in Dixon and Massey (1969). Appendix B, Table C, presents frequently used Q_α values for $K = 2$ to 100.

In this example, the .05 probability level has been chosen and the critical range is 23. Table 4.2 presents a matrix of rank differences for the data for Table 4.1, in which the significant values have been starred. Calculation of a simple scalability index (SI) is based on dividing the number of significantly different pairs by the number of possible pairs ($k(k-1)/2$).

Table 4.2
Matrix of Rank Differences

	100	A	GS	Ex
R_i	52	49	24	19
52	-			
49	3	-		
24	*28	*25	-	
19	*33	*30	5	-
*Significant at the .05 level (critical range = 23)				

$$\begin{aligned}
 \text{Scalability Index} &= \text{No. Sig. Different Pairs} / K(K-1)/2 \\
 \text{SI} &= 4 / 6 = .67
 \end{aligned}$$

A relative scalability index may also be calculated. Given a limited number of judges the number of possible significant pairs is determined by using N times the integers 1 to K as the rank sums and testing how many of those paired sums are significantly different. These sums represent the maximum possible scale scores. Note that when the number of judges (N) exceeds the critical range [$Q_\alpha E(S)$], the relative scalability index (RSI) equals the scalability index.

$$\begin{aligned}
 \text{Relative Scalability Index} &= \text{No. Sig. Different Pairs} / \text{No. Possible Sig. Different Pairs} \\
 \text{RSI} &= 4/6 \\
 &= .67
 \end{aligned}$$

Number of Judges

If tests of significance are considered to be an important part of the scaling process then a sufficient number of judges (N) need to be selected. This will ensure that the objects have the opportunity to be significantly different. In Table 4.3 are the sample sizes necessary at

the .01, .05, and .10 probability levels. For the (K) and alpha levels not listed a solution can be obtained by solving for N in the formula at the bottom of Table 4.3 where Q is taken from the values by Harter (1959) or Dixon and Massey (1969).

Table 4.3
Number of Judges Necessary to Insure the Possibility
of K Items Being Significantly Different

Number of Judges Needed at Selected Alpha Levels			
<u>Items</u>	<u>.01</u>	<u>.05</u>	<u>.10</u>
3	17	11	9
4	33	19	18
5	53	38	31
6	80	57	47
7	112	82	68
8	150	111	92
9	194	145	123
10	244	184	157
11	301	228	195
12	364	278	239
13	434	333	287
14	511	394	341
15	594	461	400
$N = Q^2 (K)(K + 1) / 12$			

Discussion

Multiple comparison tests of significance would appear to be a useful adjunct to the scaling process for the following reasons:

1. Tests of significance can provide help in making decisions about whether two objects come from the same population of stimuli. In the example presented, **100** and **A** were scaled very close to each other. The difference between the sums for **A** and **100** is three. The small difference of three or less has an extremely high probability of occurrence by chance so that the observed difference is likely just a chance difference. Thus **100** and **A** (which can be categorized as adult approval rewards) could be chosen to represent one category of reward preference on alternate forms of an instrument that compares adult approval with independence rewards. That is, a child might be asked whether he or she preferred an "A" to "being free to do what he or she liked " on one form and whether he or she preferred a "100" to "being free to go outside" on another (Dunn-Rankin & King, 1969).
2. Significance tests can create categories of stimuli that may be considered discrete. In Likert scaling, for example, the categorical descriptions can be tested to see that they were significantly different as well as equally spaced.

3. Significance tests can be used to build an index of scalability for psychological objects somewhat analogous to Guttman's coefficient of reproducibility. One index consists of a ratio of the number of significantly different pairs of objects to the total number of possible pairs. In the illustration provided in this chapter (see bottom of Table 4.2) there were four significantly different pairs out of six. This resulted in a scalability index of .67.

A relative scalability index (Knezek, 1979) can also be calculated by forming the ratio of the significantly different pairs over the number that **could be** significantly different. The denominator is found by first using the $N(K - 1)$ values as the maximum possible scale scores, then testing these pairs of values for significance, and finally counting the number of significant pairs. Such indices can be utilized to quantify the ability of different groups of people to distinguish between psychological objects.

4. By solving for N , the number of judges necessary to insure that all objects have a possibility of being significantly different at the .05 alpha level can be determined.

To summarize, this method of scaling has value because:

1. It is simple and easy to use.
2. It allows scaling along a continuum with meaningful end points.
3. Its scale values correlate highly with those obtained by other techniques.
4. It allows for tests of significance between the items.

The scale scores obtained by this simplified rank method can be utilized in traditional ways and are strikingly isomorphic (a close one-to-one relationship) with values obtained under Thurstone's Case V Model (Thurstone, 1927).

Smith (1968) showed, in a study of 19 actual paired comparison matrices varying in size from four to 21 items, a decidedly linear distribution regardless of the number of items in the matrix. The correlation between the actual distribution and the theoretical uniform distribution varied between .967 and 1.00. This indicates that purposeful judgments tend to create uniform distributions of responses. Related techniques are also proposed by Mosteller (1958), Guilford (1954) and Rummel (1964).

Application 1: Direct Ranking of Counselor Roles

Furlong, Atkinson, and Janoff (1980) developed a list of 14 elementary school counselor roles such as program planning, counseling, pupil assessment, and disciplinarian. Fifty-four counselors ranked the list in two ways: (1) the order in which they **actually** spend their time; and (2) the amount of time they **would like** to spend practicing each of the roles. Figure 4.2 shows the results of this scaling. By visually comparing ideal and actual scale scores it can be seen that the role of disciplinarian has the greatest disparity. The authors concluded that research and evaluation, (which comprises a role at the low end of the scale) is the least preferred (both actual and ideal) activity for elementary counselors.

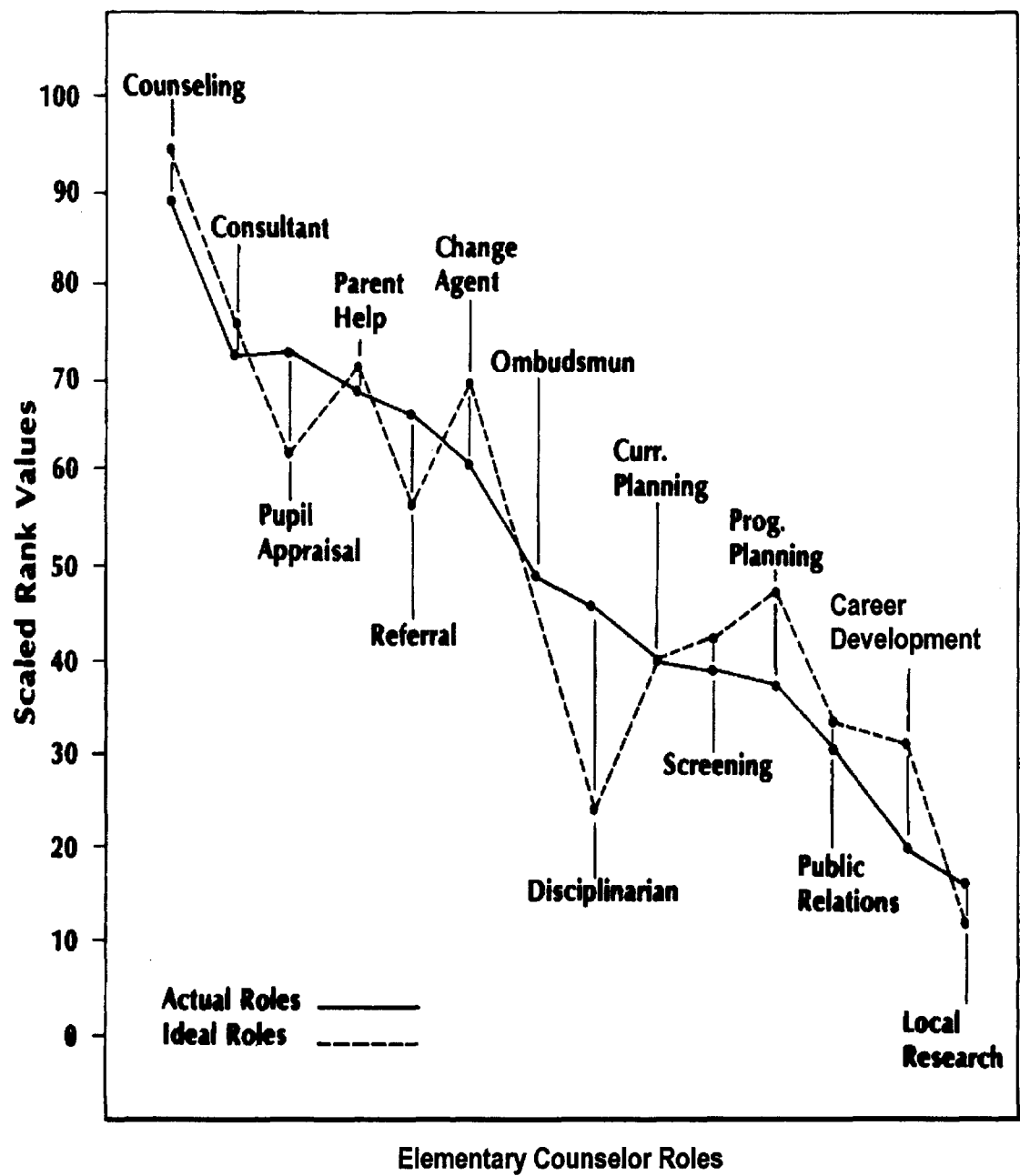


FIG. 4.2. Difference between actual and ideal roles for 54 counselors. The largest difference between roles occurs for disciplinarian.

Application 2: Letter Similarity Scales

Dunn-Rankin (1968, 1978) scaled lowercase letters of the English alphabet using the Simplified Rank Method (see Figs. 4.3 and 4.4). In this study, the objects (Century School Book font lower case letters) were paired and votes (rank values) obtained. Each child was asked to circle one letter of each pair that looks most like the target on the left. Figure 4.3 illustrates one child's task.

Name _____

Sex _____ Age _____ Grade _____

In every block be sure you circle
one letter of each pair that is most
 like the single letter.
 DO NOT SKIP ANY PAIRS.

e	a	b	h	g	t	r	o	m	l	r	m	d	t	m	m	w	m
	y	c		h	s		n	a		p	n		s	n		u	n
	w	d		i	r		d	c		o	a		r	o		t	o
	u	e		k	p		e	b		e	d		p	a		s	p
	t	f		l	o		f	y		f	c		f	e		r	a
s	s	g	n	m	n	c	g	w	f	g	b	k	g	d	y	g	f
	r	h		a	c		h	u		h	y		h	c		h	e
	p	i		b	d		i	t		i	w		i	b		i	d
	o	k		y	e		k	s		k	u		k	y		k	c
	n	l		w	f		l	r		l	t		l	w		l	b
g	m	a	i	u	g	p	m	p	b	m	s	u	m	u	t	m	y
	c	b		t	h		n	o		n	r		n	t		n	w
	d	y		s	i		a	d		o	p		o	s		o	u
	e	w		r	k		c	e		a	e		p	r		p	t
	f	u		p	l		b	f		d	f		a	f		r	s

FIG. 4.3. Letter similarity instrument, one of 21 pages of pairs.

Letter Similarity Scales shown in Fig. 4.4 were constructed (Dunn-Rankin, 1968) on the basis of a letter similarity test. In the test instrument, 21 lowercase letters were matched in pairs to determine their relative similarity to one another. The cumulative choices made by 315 second and third-grade children over all 210 possible pairings were then analyzed. Sets of linear scale values were assigned to the letters in terms of their relative similarity to each target letter. (The comparatively infrequent letters not included in the study were j q v x and z because to include them made the task too time consuming for young children.) In the arbitrary frame of reference in which the letters are plotted, zero indicates no differences among the judges in choosing a letter most like the target letter. 100 indicates no differences in choosing the letter most unlike the target letter. The scales are organized by similar letter groups

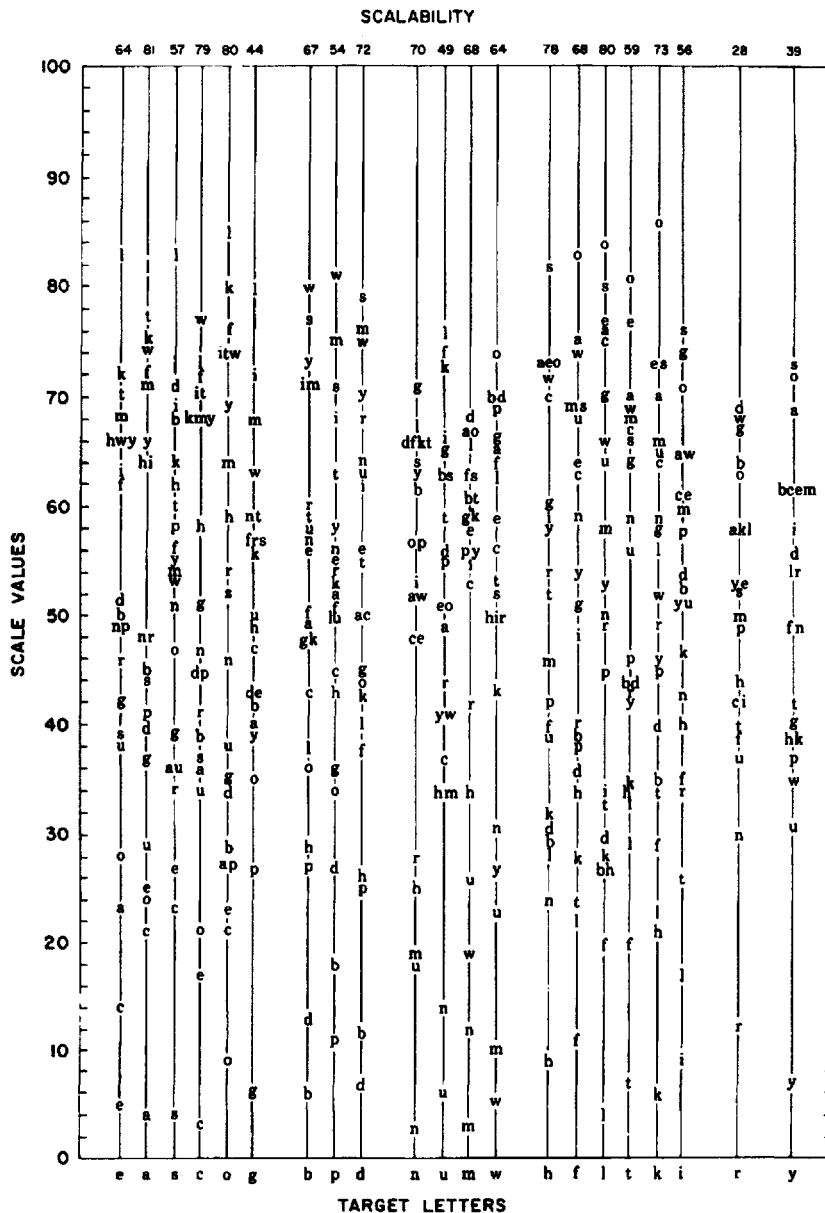


FIG. 4.4. Letter similarity scales.

CD-ROM Example Using RANKO

Available on the CD-ROM is the variance stable scaling program called RANKO. RANKO can handle ranked data as well as paired comparisons. Using the Data from the example presented in Table 4.1 and presenting the data as ranks in reverse order the analysis is as follows:

reward.cfg

```
Reward Preferences (data switch=1)
24 4 1
reward.dat
reward.out
```

Subjects, Objects, Switch. 1 means rank order data

reward.dat

```
2 1 3 4
1 2 4 3
1 2 3 4
2 2 2 4
4 3 1 2

1 2 3 4
2 1 4 3
1 4 3 2
1 2 4 3
3 1 3 3

2 1 3 4
2 1 4 3
1 2 4 3
4 3 1 2
2 1 3 4

1 2 4 3
4 1 3 2
1 2 3 4
1 4 3 2
1 4 2 3

3 1 3 3
1 2 4 3
1 2 3 4
1 2 3 4
```

Data in rank form

reward.out

Reward Preferences (data switch = 1)

Scale scores from ranks

Subjects = 24

Categories = 4

Data Type = 1

Subject	Ranks
1	2 1 3 4
2	1 2 4 3
3	1 2 3 4
4	2 2 2 4
5	4 3 1 2
6	1 2 3 4
7	2 1 4 3
8	1 4 3 2
9	1 2 4 3
10	3 1 3 3
11	2 1 3 4
12	2 1 4 3
13	1 2 4 3
14	4 3 1 2
15	2 1 3 4
16	1 2 4 3
17	4 1 3 2
18	1 2 3 4
19	1 4 3 2
20	1 4 2 3
21	3 1 3 3
22	1 2 4 3
23	1 2 3 4
24	1 2 3 4

Item	Rank Total	Scale Score
------	------------	-------------

Min	24	0
1	43	26
2	48	33
3	73	68
4	76	72
Max	96	100

0 10 20 30 40 50 60 70 80 90 100

+---+---+*+*+---+---+*+*+---+---+

Table of Rank Differences

	Items			
	4	3	2	1
4	0			
3	3	0		
2	28	25	0	
1	33	30	5	0

Significance Level	Rank Difference Needed
--------------------	------------------------

0.999	1
0.995	2
0.990	3
0.975	4
0.050	10
0.025	10
0.010	11
0.005	11
0.001	12

Circular Triad Analysis

Circular triads are proposed as a basis for determining object scalability and individual judge consistency when using complete paired comparisons data. Circular triads are formed whenever intransitive (inconsistent) pair wise choices occur. If a subject is presented with all possible pairs of three objects, A, B, and C, for example, and is asked on each occasion to judge which member of the pair is preferred, then a preference pattern of the following type may result:

$$A > B, B > C, C > A \quad \text{where } > \text{ means "is chosen over."}$$

This preference pattern is called a circular triad. The first object is preferred to the next, whereas the remaining object is preferred to the first. There is no indication of which object is most or least preferred. Whenever a circular triad exists, a nonlinear ordering has taken place. When **no** circular triads occur in a set of paired data a linear ordering results.

Because all **intransitive** preference patterns, involving more than three objects, can be decomposed into circular triads (Kendall & Babington-Smith, 1939), the number of circular triads can serve as an index of intransitivity in complete paired comparisons data.

Judge Circular Triads (JCT)

Kendall and Babington-Smith (1939) suggested that the relative consistencies of judges could be determined by counting the total number of circular triads each judge produced in the course of making choices among all possible pairs of objects. They were unable to determine **exactly** how inconsistent any **one** judge was because the exact probability distributions of Circular Triads for more than eight objects were unknown. Knezek, Wallace and Dunn-Rankin (1998) were able to solve this problem by using a computer to generate the true distributions from 5-15 objects. These exact distributions are given in Table E in Appendix B. The authors were also able to show that the Chi Squared approximation is a very accurate approximation to the true distribution. For k objects larger than 15 see one of the two references above.

Kendall (1955) developed a method for calculating the number of judge circular triads (JCT) from a vector containing a judge's preference for each object.

$$JCT_i = K(K-1)(2K-1) / 12 - \sum a_{ij}^2 / 2$$

Suppose there are $k = 5$ objects A, B, C, D, E and a judge's vote vector showing ties indicating circular triads is 2, 2, 3, 1, 2. a_{ij} = number of times judge_i preferred object_j. The votes for each object are squared and summed. $\sum a_{ij}^2 = 22$.

$$\begin{aligned} JCT &= 5(4)(9) / 12 - 22 / 2 \\ &= 4 \text{ circular triads.} \end{aligned}$$

The four circular triads are (1) $A > B > C > A$; (2) $A > D > E > A$; (3) $B > C > D > B$ and (4) $B > E > D > B$

Coefficient of Consistency

Kendall and Babington-Smith (Kendall, 1955) developed measures of judge consistency also based on the number of circular triads. These formulas are:

$$\begin{aligned} \text{Consistence}_i &= 1 - 24(\text{JCT}_i)/(K^3 - K) && \text{for odd } K, \text{ and} \\ \text{Consistence}_j &= 1 - 24(\text{JCT}_j)/(K^3 - 4K) && \text{for even } K \end{aligned}$$

Where K = number of objects and JCT_i is the number of circular triads produced by Judge $_i$.

Tests for Circularity

Suppose one conducts a complete paired comparison experiment and is interested in analyzing the confusions that occur. One might ask "How consistent were the judges?; What objects did they involve in circularity? What objects should be removed from the final scale?"

1. **Judge Consistency:** One can test whether an individual **judge** has significantly fewer circular triads, across all the objects, than is expected by chance. The researcher can use the values in Table E as a check.

2. **Overall Circularity:** One can test whether the **total** circularity that occurs for all the judges across all the objects is less than one would expect by chance (Knezek, 1979). This test should generally be significant at the conventional alpha level (.05) because if the data are discriminable and the dimension of judgment well defined, then logical, not chance, choices will prevail. Consistency however, can be determined using other p levels, $p < 0.20$ for example.

3. **Relative Consistency:** One can test whether an individual **judge** has a significantly greater number of circular triads than the group average using a Z test.

4. **Object Circularity:** One may wish to know whether a particular **object** is involved in significantly less circularity across all the judges than would be expected by chance. This test will normally not be significant at the conventional level of .05. A conservative probability can found using Cheybshev's inequality (Gnedenko & Khinchin, 1962): OCT indicates the Object Circular Triad involvement.

$$P\{|\text{OCT} - \text{Mean}_{\text{OCT}}| > d\} < \text{Variance}_{\text{OCT}} / d^2 \quad \text{where}$$

d is chosen by the experimenter, OCT = object circular triads, K = the number of objects, $\text{Mean}_{\text{OCT}} = \{(K^2 - 3K + 2)/8\} / N$ and $\text{Variance}_{\text{OCT}} = 3/4 \text{Mean}_{\text{OCT}}$.

5. **Relative Object Circularity:** One may wish to test, therefore, whether a specific **object** is involved in more or less circularity than other objects, that is, a relative test. In this case, the Z statistic can be used after calculating (using conventional formulas) the mean and standard deviation of object circular triads for a particular set of data.

6. Pairwise Object Circularity: Circular triads are sometimes associated with a particular pair of objects, and these circularities are often unidirectional. Therefore, one may wish to test whether the **direction** of the preferences between two objects, involved in a large amount of circularity, is significantly greater than expected by chance. Here the traditional binomial test can be employed: where X runs from S to T (Gnedenko & Khinchin, 1962):

$$P(X > S) = \sum (X \text{ choose } T) p^X (1 - p)^{T - X}$$

where T = the number of times a particular pair of objects is involved in a circular triad (the number of trials), S = the larger number of single direction occurrences (the number of successes), p = .50 is the random probability of each preference being in the direction with the larger number of occurrences (the probability of success on each trial), and X is a random variable that can take on any value within the range of the distribution.

Because the value yielded is a one-tailed probability, it should be doubled whenever a specific directionality is not hypothesized. A computer program is usually necessary to count directional circularity. This is provided in program **TRICIR** on the CD-ROM.

7. Coefficient of Variation: It is helpful to calculate the coefficient of variation $CV = [100 (\text{Standard Deviation}/\text{Mean})]$ among object circular triads for each pair of objects involved in circularity. This determines whether circularity is evenly distributed across the judges. A large coefficient indicates that only a few judges are involved in the circularity for that particular pair of objects, whereas a small coefficient shows the circularity is widely distributed.

Application: Circularity Among Adjective Pairs

An example serves to illustrate some of the procedures described in the foregoing section. In a study (Dunn-Rankin, Knezek, & Abalos (1978), 15 adjectives were selected on the criterion that all were socially desirable. All adjective pairs were presented and 39 high school students were asked to choose the trait preferred in each case (Fig. 4.5 illustrates the instrument used). The resulting rank ordering and scale values were based on rank scaling.

DIRECTIONS

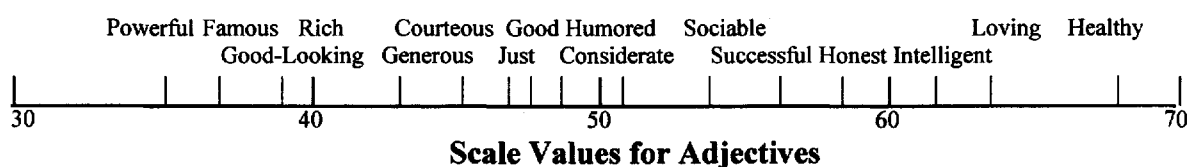
You will find 105 pairs of words below. For each pair of words, you are to choose one according to your own preference. Underline your choice for each pair of words. For example: Healthy or Sociable. If you prefer to be sociable rather than healthy, then underline the word Sociable. There is no time limit for this task, so take your time!

WHAT WOULD YOU RATHER BE?

- | | |
|-----------------------------|--------------------------------|
| 1) Good-Looking or Sociable | 54) Loving or Healthy |
| 2) Healthy or Good-Looking | 55) Rich or Generous |
| 3) Generous or Just | 56) Just or Loving |
| 4) Honest or Famous | 57) Sociable or Courteous |
| 5) Healthy or Sociable | 58) Sociable or Successful |
| 6) Generous or Courteous | 59) Good-Humored or Powerful |
| 7) Successful or Just | 60) Good-Looking or Just |
| 8) Powerful or Good-Looking | 61) Considerate or Intelligent |
| 9) Loving or Courteous | 62) Honest or Generous |
| 10) Loving or Generous | 63) Sociable or Honest |
| 11) Just or Rich | 64) Sociable or Loving |

12)	Healthy or Successful	65)	Courteous or Considerate
13)	Healthy or Honest	66)	Successful or Good-Looking
14)	Good-Humored or Loving	67)	Rich or Courteous
15)	Powerful or Loving	68)	Good-Humored or Just
16)	Successful or Generous	69)	Intelligent or Generous
17)	Good-Humored or Famous	70)	Famous or Loving
18)	Considerate or Honest	71)	Powerful or Sociable
19)	Intelligent or Healthy	72)	Loving or Rich
20)	Considerate or Healthy	73)	Generous or Healthy
21)	Healthy or Powerful	74)	Generous or Famous
22)	Powerful or Courteous	75)	Honest or Good-Looking
23)	Honest or Good-Humored	76)	Considerate or Powerful
24)	Powerful or Intelligent	77)	Generous or Powerful
25)	Intelligent or Loving	78)	Good-Humored or Good-Looking
26)	Healthy or Rich	79)	Good-Looking or Loving
27)	Just or Considerate	80)	Good-Looking or Considerate
28)	Healthy or Courteous	81)	Courteous or Good-Humored
29)	Intelligent or Famous	82)	Courteous or Good Humored
30)	Intelligent or Sociable	83)	Successful or Good-Humored
31)	Loving or Intelligent	84)	Good-Looking or Generous
32)	Rich or Successful	85)	Powerful or Honest
33)	Famous or Considerate	86)	Honest or Successful
34)	Rich or Good-Humored	87)	Powerful or Successful
35)	Good-Humored or Sociable	88)	Good-Looking or Intelligent
36)	Rich or Powerful	89)	Healthy or Just
37)	Considerate or Just	90)	Loving or Successful
38)	Just or Powerful	91)	Generous or Sociable
39)	Good-Humored or Considerate	92)	Loving or Considerate
40)	Good-Humored or Considerate	93)	Just or Sociable
41)	Famous or Rich	94)	Sociable or Considerate
42)	Loving or Honest	95)	Rich or Good-Looking
43)	Rich or Honest	96)	Famous or Healthy
44)	Considerate or Successful	97)	Intelligent or Honest
45)	Intelligent or Courteous	98)	Courteous or Honest
46)	Rich or Intelligent	99)	Famous or Sociable
47)	Famous or Powerful	100)	Just or Honest
48)	Famous or Just	101)	Courteous or Good-Looking
49)	Good-Humored or Intelligent	102)	Successful or Intelligent
50)	Considerate or Generous	103)	Healthy or Good-Humored
51)	Successful or Famous	104)	Good-Looking or Famous
52)	Courteous or Just	105)	Rich or Sociable
53)	Successful or Courteous		

FIG. 4.5. 105 pairs of adjectives presented to 39 high school students.



Circular Triad Analysis

1. Judge Consistency. Most of the students were highly consistent in their choices. There was, however, one exception. One student produced 107 circular triads. Because the number produced is greater than the $p = .05$ critical value of 97 (probability = .053) shown in Table E of Appendix B, (see p. 220) the null hypothesis of random choices is retained for this judge. One can assume this judge was guessing or doubt the judge's competence. In either case, the student's data should probably be removed from the analysis.

2. Overall Circularity. Although words, thought to be close to each other in social desirability, were chosen with hopes that high circularity would be produced, the adjectives instead proved to be highly scalable. Only 1343 judge circular triads were produced by the 39 judges, an average of $(1343/39)$ or 34. This is far fewer than the average of 113.75 expected by chance $[15(15 - 1) / (15 - 2)/24]$. The Coefficient of Consistence, $C = 1.0 - 24(34.43)/5360$ is high at 0.754. Note that the student with 107 circular triads is very close to the chance average of 113.75.

3. Relative Judge Consistency. This measure is calculated by finding the number of circular triads for all judges. Then determine the Mean and Standard Deviation and calculate Z to identify outliers. In this area, a strong case can again be made against the aforementioned student. The relative reliability of two other students can also be questioned. The first student's 107 circular triads, standardized utilizing the Mean and Standard Deviation of all JCTs, is transformed to a Z score of +3.1. The other two judges each produced 88 circular triads, each equivalent to a Z score of +2.29. All these scores are significantly inconsistent at the .05 level, according to the group distribution under the assumption of the applicability of the standard normal curve. (A Kolmogorov-Smirnoff test uncovered no deviation $[p = .05]$ from normality in the data). It is possible therefore to remove the choices of all three students from the analysis.

4. Absolute Object Scalability. The object involved in the greatest number of circular triads (sociable) was subjected to this test. If $NK > 16$, the distribution of the Object Circular Triads (OCT) will be approximately normal, therefore, a Z test was used. The 370 object circular triads for "sociable" were equivalent to a theoretical Z score of -3.21, with a probability level of less than .001, where

$$\begin{aligned} Z &= [\text{OCT}_k - (N)(\text{Mean}_{\text{OCT}})] / (N)(\text{SD}_{\text{OCT}}) \\ &= [370 - (39)(22.75)] / (39)(4.13) \\ &= -3.21. \end{aligned}$$

(as Z values approach zero, objects are less scalable). The null hypothesis of random scalability would therefore be rejected. Because **sociable** was found to be scalable, all other objects, that have fewer circular triads, would also be highly scalable.

5. Relative Object Scalability. Three objects were selected for this test. They were the two with the highest number of circular triads and the one with the lowest. Using the mean and standard deviation of the entire group of object circular triads, Z scores of + 2.28, + 1.24, and - 1.83 were derived for **sociable**, **intelligent**, and **powerful**, respectively. Sociable deviates significantly ($p = .05$) from the circularity of others in the group, whereas the remaining two objects do not. If the goal is improvement of the scale of social desirability, the word sociable would be eliminated from the scale.

6. Pairwise Circularity. A large number of circular triads (101) were found to be associated with the pair of objects **sociable** and **rich**. Because this number was much larger than the average value of 38.37, for all pairs involved in circular triads, and because 80 of these 101 involvements were in the direction of **sociable** being preferred to **rich** (binomial probability < 0.01), the pair was singled out for further study. The word **rich** was found to be involved in only

about the average (38) number of circular triads, whereas **sociable**, as previously stated, was involved in a significantly greater amount of circularity than the remainder of the objects in the group. These facts, in combination with consideration of a low coefficient of variation for **sociable** (1.22 versus a mean c.v. of 1.49), illustrate that the word **sociable** is often preferred to **rich**, **rich** is preferred to many other words, and these words are then preferred to **sociable**. Removal of the word **sociable** should eliminate this inconsistent adjective, leading to an improvement in a linear scale of social desirability.

Discussion

The analysis of data through circular triads should be especially appropriate whenever the concept of judge reliability does not depend on interjudge agreement. Data from paired comparison instruments such as the Edwards Personal Preference Schedule (EPPS, Edwards, 1959) might meet this requirement. Circular triad analysis of such data would determine the degree to which individuals were consistent in their personality choices, as well as test for the overall consistency of the group. The test of absolute judge consistency allows for the possibility of perfect consistency within a judge even if the ordering of the objects for every judge is unique.

The exact cumulative probabilities for circular triads, given in Table E, Appendix B, are limited to the upper and lower 10% tails. For other values and K larger than 15 one can use Kendall's chi-square approximation (Kendall & Babington-Smith, 1939):

$$\chi^2 = \frac{8}{(k-4)} \left\{ \frac{1}{4} (k \text{ choose } 3) - d - \frac{1}{2} \right\} + v \quad v = \frac{k(k-1)(k-2)}{(k-4)^2}$$

where k is the number of objects and d is an individual's number of circular triads. For example if a person had 13 circular triads, given 10 objects, $\chi^2 = 42$ and $v = 20$. Knezek, Wallace, and Dunn-Rankin (1998) showed that the Chi-square estimate of the probability for 13 or fewer circular triads is .0028, close to the actual cumulative probability value the authors enumerated to be .0021. To aid in the analysis of circular triads, the computer program, **TRICIR**, is available on the CD-ROM.

CD-ROM Example Using TRICIR

The disability data from Chapter 1 (p. 5) demonstrates the use of the **TRICIR** program. In this brief example some tests for circularity are inappropriate because the number of objects is too small ($n = 4$). First a configuration file is constructed as a text file. See **tricir_readme.txt** for a full explanation. The file consists of:

1. title line;
2. a parameter line (the number of judges, number of variables, 1 for full output, 0 for a summary, data code (0 = 1 and 0, 1 = 1 and 2), separated by spaces;
3. the input file name. The input file contains the key code for the pairings, each code separated by blanks. This is followed by the pair choices, a 1 for the first of each pair or a 2 recorded if the second of each pair is selected; and
4. the output file name.

disabltty.cfg (Configuration file)

Disability Data Set			
5	4	0	1
disab.dat			
disab.out			

Title
 # Judges, # Objects, 0 = full output, 1 = 1-2 coding
 Input file Name
 Output file Name

disab.dat (Input File)

01 02 01 03 01 04 02 03 02 04 03 04
1 1 1 2 2 2
1 1 1 2 2 2
1 1 1 1 2 2
1 1 2 2 2 2
1 1 1 1 2 1

Pairing Key

Pair Choices (1 and 2)

In this analysis Object 1 = Learning Disability, 2 = Mentally Retarded, 3 = Deaf, and 4 = Blind

disab.out (Output File)

Disability Data Set

Analysis summary

Object	# CT's In	ABS Z	ABS Prob	Grp Z	# Votes	Scaled
1	0	-1.00	0.1587	-1.50	14.	93.33
2	1	-0.73	0.2317	0.50	2.	13.33
3	1	-0.73	0.2317	0.50	4.	26.67
4	1	-0.73	0.2317	0.50	10.	66.67

Kendall's Coefficient of Concordance (W) for Judges Votes**W = .7913 Prob (x>=W) = .0078**

Probability not accurate for 7 or fewer objects (small numbers of objects are not normally distributed)

Judge	# CT	Consis	Abs Prob	Grp Z
1	0.	1.0000	0.1968	-0.45
2	0.	1.0000	0.1968	-0.45
3	0.	1.0000	0.1968	-0.45
4	0.	1.0000	0.1968	-0.45
5	1.	0.5000	0.4468	1.79

Mean # CT's = 0.200

Standard deviation = 0.447

Average consistency = 0.9000

Prob(<= 1. ct's for 5 Judges) = .0004000

Significant Scale Differences for Objects

Objects	Diff	P=.1	P=.05	P=.01
1 2	80.00	1.	1.	0.
1 3	66.67	1.	0.	0.

Critical Differences 63.37 70.83 85.47
 (Approximate)

May not be accurate for 15 or fewer judges (See RANKO for table of exact values)

Significantly different pairs ($p=.05$) = 1.

Possible significant pairs ($p=.05$) = 1.

Relative scalability index = 1.0000

Probability ($x \geq \text{index}$) = 0.0500

Judge: 1

Object	# CT's in	ABS Z	ABS Prob	# Votes	Scaled
1	0	-1.00	0.1587	3.	100.00
2	0	-1.00	0.1587	0.	0.00
3	0	-1.00	0.1587	1.	33.33
4	0	-1.00	0.1587	2.	66.67

Total circular triads for judge 1 = 0.

Kendall's coefficient of consistence = 1.0000

$\text{prob}(x \leq \#ct) = \text{prob}(x > \text{coef}) = .196784884$

Judge: 2

Object	# CT's in	ABS Z	ABS Prob	# Votes	Scaled
1	0	-1.00	0.1587	3.	100.00
2	0	-1.00	0.1587	0.	0.00
3	0	-1.00	0.1587	1.	33.33
4	0	-1.00	0.1587	2.	66.67

Total circular triads for judge 2 = 0.

Kendall's coefficient of consistence = 1.0000

$\text{prob}(x \leq \#ct) = \text{prob}(x > \text{coef}) = .196784884$

Judge: 3

Object	# CT's in	ABS Z	ABS Prob	# Votes	Scaled
1	0	-1.00	0.1587	3.	100.00
2	0	-1.00	0.1587	1.	33.33
3	0	-1.00	0.1587	0.	0.00
4	0	-1.00	0.1587	2.	66.67

Total circular triads for judge 3 = 0.

Kendall's coefficient of consistence = 1.0000

$\text{prob}(x \leq \#ct) = \text{prob}(x > \text{coef}) = .196784884$

Judge: 4

Object	# CT's in	ABS Z	ABS Prob	# Votes	Scaled
1	0	-1.00	0.1587	2.	66.67
2	0	-1.00	0.1587	0.	0.00
3	0	-1.00	0.1587	1.	33.33
4	0	-1.00	0.1587	3.	100.00

Total circular triads for judge 4 = 0.

Kendall's coefficient of consistence = 1.0000

$\text{prob}(x \leq \#ct) = \text{prob}(x > \text{coef}) = .196784884$

Judge: 5

Object	# CT's in	ABS Z	ABS Prob	# Votes	Scaled
1	0	-1.00	0.1587	3.	100.00
2	1	0.33	0.6306	1.	33.33
3	1	0.33	0.6306	1.	33.33
4	1	0.33	0.6306	1.	33.33

Total circular triads for judge 5 = 1.

Kendall's coefficient of consistence = 0.5000

prob(x<=#ct) = prob(x>coef) = .446761280

Circular triad distribution (across all judges)

Objects	>	<	Total	
2 3 4	1.	0.	1.	.5000

(The > is read 1 circular triad and the direction is 2 > 3 > 4 > 2)

Total circular triads across 5 judges = 1.

Preference Matrix for Circular Triads

Objects	# A>B	# A<B	Total	Prob(x<=min)	Varcoeff
1 2	0.	0.	0.	1.0000	*****
1 3	0.	0.	0.	1.0000	*****
1 4	0.	0.	0.	1.0000	*****
2 3	1.	0.	1.	0.5000	2.2361
2 4	0.	1.	1.	0.5000	2.2361
3 4	1.	0.	1.	0.5000	2.2361
Means	0.33	0.17	0.50	0.7500	*****

Note that in this brief data there is only one circular triad. It occurs with Judge 5 and involves objects 2, 3, and 4. The judges were consistent. **TRICIR** also performs the rank scaling of the votes and these are the same as presented in the original calculation. Object 1 (LD, Learning Disability) is seen as significantly ($p < .05$) more preferred to be mainstreamed than 2 (MR, Mentally Retarded) or 3 (D, Deaf).

Kendall's coefficient of concordance, W , is a measure of agreement among raters. It can range from 0 to 1 (Siegel, 1956). The rank sums are found. If there is complete agreement among the judges the rank sums would be some multiple of the number of judges and produce the largest variability. A ratio is formed between the rank sum variability that does occur divided by the maximum variability. For the vote data of the example: (0, 5, 10, 15) $SS_{max} = 125$ and (14, 2, 4, 10) $SS_{obtained} = 91$. $W = 91/125$ or $W = .73$ (uncorrected for ties). To correct for ties subtract the $[(\text{number of judges}) \Sigma(t^3 - t)/12]$ from the denominator. t is the number of tied ranks. Judge 5 has three tied ranks, $t = 3$. Subtract $5(27 - 3)/12$ or 10. $W = 91/115 = .7913$.

5

ORDER ANALYSIS

Guttman Scaling

Guttman (1944, 1950) described a unidimensional scale as one in which the respondents' responses to the objects would place individuals in perfect order. Ideally, persons who answer several questions favorably all have higher skills than persons who answer the same questions unfavorably. Arithmetic questions make good examples of this type of scale. Suppose elementary school children are given the following addition problems:

$$\begin{array}{llllll}
 (1) & 2 & (2) & 12 & (3) & 28 & (4) & 86 & (5) & 228 \\
 & +3 & & +15 & & +24 & & +88 & & +894
 \end{array}$$

It is probable that if subject **A** responds correctly to item 5, he or she would also respond correctly to items 1, 2, 3, and 4. If subject **B** can answer item 2 and not item 3, it is probable that he or she can answer item 1 correctly but would be unable to answer item 4 and 5. By scoring 1 for each correct answer and 0 otherwise, a profile of responses can be obtained. If the arithmetic questions form a perfect scale, then the sum of the correct responses to the five items can be used to reveal a person's scale type. Their sum, is reflected in a series of ones and zeros. In our example:

	Items					
	1	2	3	4	5	Sum
Subject A has scale type	1	1	1	1	1	= 5
Subject B has scale type	1	1	0	0	0	= 2

Given a perfect scale, the single summed score reveals the scale type. Thus a single digit can be used to represent all the responses of a person to a set of items. Guttman Scales have been labeled deterministic. With five questions and scoring the item as correct or incorrect there are only six possible scale types. These are:

1	2	3	4	5	6
1 1 1 1 1	1 1 1 1 0	1 1 1 0 0	1 1 0 0 0	1 0 0 0 0	0 0 0 0 0

Although there exist 32 possible arrangements (2^5) of five ones and zeros, only six of these form scale types. In general, the **number of scale types for dichotomously** scored data is $(K + 1)$, where K is the number of objects. Although a perfect Guttman scale is unlikely to be found in practice, approximations to it can be obtained by a judicious choice of items and careful analysis of a set of responses to a larger number of items than are to be used in the final scale.

Goodenough’s Error Counting

Goodenough revised error counting in Guttman Scaling in a method sometimes known as Scalogram Analysis (Edwards, 1957). First, a set of psychological objects is selected. The objects should be ones that will differentiate participants with varying attitudes or perceptions about the objects along some **single** dimension. Suppose the following six statements have been chosen and 12 student’s responses have been obtained in the form of Agreement or Disagreement with these statements. Do these statements constitute a Guttman Scale along the dimension of attitudes toward school? The students are scored (1) for agree and (0) for disagree with the statements. Results are presented in Table 5.1.

Statements	Agree	Disagree
A. School is OK.	_____	_____
B. I come to school regularly.	_____	_____
C. I think school is important.	_____	_____
D. It is nice to be in school.	_____	_____
E. I think school is fun.	_____	_____
F. I think school is better than a circus.	_____	_____

Table 5.1

Item Response Data in the Form of Ones and Zeros							
Students	A	B	C	D	E	F	Scores
1	0	1	1	1	1	0	4
2	1	1	1	0	0	0	3
3	1	0	0	0	0	1	2
4	1	1	0	0	0	0	2
5	0	0	1	1	1	0	3
6	0	1	0	1	1	0	3
7	0	1	0	0	1	0	2
8	0	1	1	0	0	0	2
9	1	1	0	1	1	1	5
10	0	1	1	1	0	0	3
11	0	0	1	0	0	0	1
12	0	1	1	0	1	0	3
Sum	4	9	7	5	6	2	33

The basic data are arranged in a table of ones and zeros in which **one (1)** stands for agree and **zero (0)** for disagree, and the rows and columns are summed.

It is initially convenient to rearrange the first table in order of the **row and column sums** (see **Table 5.2**). Errors in scale types are calculated by subtracting the profile of ones and zeros for each respondent from the perfect scale type with **the same summed** score. For example:

Perfect Scale Type for a score of 5 = 1 1 1 1 1 0
 Subject 9 has a score of 5 = 1 1 0 1 1 1
 Difference = 1 -1

The sum of the **absolute value** of each difference is the error. In this case $1 + |-1| = 2$ errors.

Table 5.2

Rearrangement of Ones and Zeros

Students	B	C	E	D	A	F	Scores	Errors
9	1	0	1	1	1	1	5	2
1	1	1	1	1	0	0	4	0
2	1	1	0	0	1	0	3	2
5	0	1	1	1	0	0	3	2
6	1	0	1	1	0	0	3	2
10	1	1	0	1	0	0	3	2
12	1	1	1	0	0	0	3	0
3	0	0	0	0	1	1	2	4
4	1	0	0	0	1	0	2	2
7	1	0	1	0	0	0	2	2
8	1	1	0	0	0	0	2	0
11	0	1	0	0	0	0	1	2
Sum	9	7	6	5	4	2	33	20

The total **possible** number of errors is equal to the product of N subjects and K objects (items) or in this case $(N)(K) = (12)(6)$ or 72 possible errors. This is because the maximum possible number of errors is six for any one subject. An estimate of how accurately the particular arrangement approximates a perfect scale is to take the ratio of the found errors to the maximum number of possible errors. Subtracting this ratio from 1.0 renders a coefficient of the scale's ability to reproduce the scores based on the row sums. The **Coefficient of Reproducibility** is :

$$CR = 1 - (20 \text{ Total Errors} / 72 \text{ Possible Errors}) = .723.$$

A **Coefficient of Scalability**, $CS = 1 - E / X$ is determined by finding the ratio of the number of errors (E) to the number of errors by chance (X), subtracted from 1. That is, $CS = 1 - \text{No. of Errors} / .5(N)(K)$ or $1 - 20 / .5(12)(6) = 0.45$. The original items form a poor approximation to a true Guttman scale. 0.60 has been suggested as the lower limit for CS. Item C (I think school is important.) accumulates six errors, out of a possible 12, so it is reasonable to eliminate this item. After eliminating item C the responses to the remaining items are reorganized. In this second reorganization (Table 5.3), 12 errors occur.

Table 5.3

Reduced Matrix of One and Zeros

Judges	B	E	D	A	F	Score	Error
9	1	1	1	1	1	5	0
1	1	1	1	0	0	3	0
6	1	1	1	0	0	3	0
12	1	1	0	0	0	2	0
7	1	1	0	0	0	2	0
2	1	0	0	1	0	2	2
4	1	0	0	1	0	2	2
10	1	0	1	0	0	2	2
5	0	1	1	0	0	2	2
3	0	0	0	1	1	2	4
8	1	0	0	0	0	1	0
11	0	0	0	0	0	0	0
Sum	9	6	5	4	2	26	12
p	.75	.50	.42	.33	.17		
q	.25	.50	.58	.67	.83		

Note that four errors occur for Judge 3. This person's data should be checked for accuracy in following instructions or for errors in coding. In this new arrangement, a coefficient of the ability of the column sums to reproduce all the responses accurately is $CR = 1 - \text{Errors} / (N)(K)$ or $1 - 12 / 60 = .80$. The CR index has been slightly improved by deleting item C from the analysis. $CS = 1 - 12 / .5(12)(5) = .60$. Because further deletion appears not to be of benefit (that is, does not increase the coefficient of reproducibility) no more rearrangements are performed.

A test of the effectiveness of a reproducibility coefficient can be made in light of the minimal reproducibility possible given the average proportion of agree (p) and disagree ($q = 1 - p$)

responses in each column of Table 5.3. By averaging the maximum of p or q in each column, the **Minimal Marginal Reproducibility (MMR)** can be obtained. Thus the sum of (.75 + .50 + .58 + .67 + .83) / 5 = .67. The difference between a **CR** of .80 and the **MMR** of .67 is .13. This number is the **Percentage of Improvement (PI)**.

It is possible to assign a coefficient of reproducibility to a given respondent. This coefficient may be obtained by subtracting the individuals's profile from a perfect scale vector with the same score. The sum of the absolute differences divided by K items and subtracted from 1 gives the **CR** for the subject chosen. For example if there are nine objects and a single respondents total score is six then:

Objects, K = 9		1	2	3	4	5	6	7	8	9	
Scale type	Y =	1	1	1	1	1	1	0	0	0	Score = 6
Subject's vector	X =	1	1	1	0	1	1	1	0	0	Score = 6
						1		-1			

$CR_i = 1 - [(1 + |-1|)/9] = 1 - 2/9 = .7778$

Application 1: Cloze Tests in Reading

F. J. King (1974) has utilized scalogram analysis (Guttman Scaling) to grade the difficulty of cloze tests. A cloze test asks the subject to complete passages in which a specific scattered set of words has been deleted. King has suggested that a system for getting children to read materials at their appropriate reading level must be capable of locating a student on a reading level continuum so that he or she can read materials at or below that level. This is what a cloze test is designed to measure.

King constructed eight cloze passages ordered in predicted reading difficulty. If a student supplied seven of 12 words on any one passage correctly, he was given a score of one. If he had fewer than seven correct answers, he received a score of zero. A child's scale score could vary from **zero** to **eight**, a score of 1 for each of the eight passages. If the test passages form a cumulative scale, then **scale scores should determine** a description of performance. A scale score of 4, for example, would have the vector

11110000

and this would indicate that a student could read text material at the fourth level and below. By constructing a table to show the percentage of students at each scale level who passed each test passage, King was able to indicate that "smoothed" scale scores were capable of producing the score vectors with considerable accuracy.

Once the scalability of the passages was determined, King related the reading difficulty of the test passages to the reading difficulty level of the educational materials in general. Thus he was able to indicate which reading material a child could comprehend under instruction.

Application 2: Arithmetic Achievement

Smith (1971) applied Guttman Scaling to the construction and validation of arithmetic achievement tests. He first broke simple addition into a list of 20 tasks and ordered these tasks according to hypothesized difficulty. Experimental forms containing four items for each task were constructed and administered to elementary school children in grades 2-6. Smith reduced his 20 tasks to nine by testing the significance of the difference between the proportion passing for each pair of items. He chose items that were different in difficulty with $\alpha < .05$ for his test. The nine tasks were scored by giving a one (1) each time three or more of the four parallel items for that task were answered correctly and a zero (0) otherwise. These items were analyzed using the Goodenough Technique. The total scale score for each subject was defined as the number of the item that preceded two successive failures (zeros). Thus a subject with the following vector:

Tasks	1	2	3	4	5	6	7	8	9
Subject's vector	1	1	1	0	1	1	0	0	0

would have a scale score of 6.

Table 5.4 shows the coefficients of reproducibility obtained by Smith on the addition tests for two schools and grades 2-6.

Table 5.4			
Obtained Coefficients of Reproducibility			
Grade	Shadeville	Sopchoppy	Both
2	.9546	.9415	.9496
3	.9603	.9573	.9589
4	.9333	.9213	.9267
5	.9444	.9557	.9497
6	.9606	.9402	.9513
All	.9514	.9430	.9487

The high coefficients of reproducibility indicate that a student's scale score accurately depicts his or her position with regard to the tasks necessary in solving addition problems. For this reason Smiths' results can be used: (1) to indicate the level of proficiency for a given student; (2) as a diagnostic tool; and (3) to indicate the logical order of instruction.

Significance of a Guttman Scale

Guttman has stated that a scale with a $CR < .90$ cannot be considered an effective approximation to a perfect scale. Further study comparing significance tests from Rank Scaling suggests that a CR of .93 approximates the .05 level of significance. Other sources suggest that the CS should be greater than .60.

It is possible to test the statistical significance of a Guttman Scale by creating a chance expected scale using a table of random numbers. Two sets of error frequencies are made for each subject. That is, the obtained or observed errors and the expected errors. The observed and expected error frequencies are tested using the Chi Square distribution with $N - 1$ degrees of freedom.

CD-ROM Example Using SCALO

SCALO, using the Goodenough Technique and restricting the data to 0 or 1 responses is provided on the CD ROM. Using the attitude toward school data from Table 5.1 with item C removed, the **output** is provided below.

scalo.out

Scalo Title Line

Number of subjects = 12 Number of variables = 5

Sub\Item	2.	4.	3.	1.	5.	Sum	Err	CS(i)
9.	1.	1.	1.	1.	1.	5.	0.	1.0000
1.	1.	1.	1.	0.	0.	3.	0.	1.0000
6.	1.	1.	1.	0.	0.	3.	0.	1.0000
2.	1.	0.	0.	1.	0.	2.	2.	0.6000
3.	0.	0.	0.	1.	1.	2.	4.	0.2000
4.	1.	0.	0.	1.	0.	2.	2.	0.6000
5.	0.	1.	1.	0.	0.	2.	2.	0.6000
7.	1.	1.	0.	0.	0.	2.	0.	1.0000
10.	1.	0.	1.	0.	0.	2.	2.	0.6000
12.	1.	1.	0.	0.	0.	2.	0.	1.0000
8.	1.	0.	0.	0.	0.	1.	0.	1.0000
11.	0.	0.	0.	0.	0.	0.	0.	1.0000
	9.	6.	5.	4.	2.	26.	12.	
p	0.75	0.50	0.42	0.33	0.17			
q	0.25	0.50	0.58	0.67	0.83			

Coefficient of reproducibility (CR) = 0.8000

Minimum marginal reproducibility (MMR) = 0.6667

Percentage of improvement (PI) = 0.1333

Coefficient of scalability(CS) = 0.6000

Mokken Scales

Mokken Scales (Mokken & Lewis, 1982) are like Guttman scales but probabilistic rather than deterministic. That is, the item difficulties are taken into account when ordering the items of the scale. A subject who answers a difficult item correctly will have a high **probability** of answering an easier item correctly. Loevinger's H statistic is used to test how well an item or set of items corresponds to Mokken's idea of scalability:

$$H = 1 - E/E_o.$$

In this case E is the probability of “errors in direction” given a Guttman Scale. These have been called Guttman errors. E_o is the probability of a Guttman error under the hypothesis of totally unrelated items. In the analysis of the responses on the right, the data are arranged in a reverse order so that a 1 represents an easier item than a zero (0). Following Guttman scaling the data are initially rearranged in ascending order of ease of response using the column sums. That is, the probability of 1 is greater than the probability of a zero. In this example, an error occurs in the fifth row where a (1 0) response is counter to expectation.

Items	A	B	C
	1	1	1
	0	1	1
	0	0	1
	0	0	1
	0	1	0
Sum	1	3	4
p	.2	.6	.8
q	.8	.4	.2

Formally, let item j be easier than item i which means that the $P(X_j) = 1 > P(X_i) = 1$ or the probability j is 1 is greater than i is 1, where 1 is a correct response.

$$H_{ij} = 1 - E/E_o$$

Where $E = P(X_i = 1 \text{ and } X_j = 0)$ and $E_o = P(X_i = 1)P(X_j = 0)$

$$\begin{aligned}
 H_{12} &= 1 - 0/5[(.2)(.4)] &= 1 - 0 / (.2)(.4) &= 1.00 \\
 H_{13} &= 1 - 0/5[(.2)(.2)] &= 1 - 0 / (.2)(.2) &= 1.00 \\
 H_{21} &= 1 - 2/5[(.6)(.8)] &= 1 - .4 / (.6)(.8) = 1 - .83 &= .17 \\
 H_{23} &= 1 - 1/5[(.6)(.2)] &= 1 - .2 / (.6)(.2) = 1 - 1.67 &= -.67 \\
 H_{31} &= 1 - 3/5[(.8)(.8)] &= 1 - .6 / (.8)(.8) = 1 - .9375 &= .0625 \\
 H_{32} &= 1 - 2/5[(.8)(.4)] &= 1 - .4 / (.8)(.4) = 1 - 1.25 &= -.25
 \end{aligned}$$

For the example above, individual item (H_i) values are obtained and averaged.

$$\begin{aligned}
 H_1 &= (1.0 + 1.0)/2 &= 1.00 \\
 H_2 &= (.17 - .67)/2 = -.5/2 &= -.25 \\
 H_3 &= (.0625 - .25)/2 = .125/2 &= -.09375 \\
 \text{Sum} &&= .65625
 \end{aligned}$$

The overall mean provides an H value for the scale. $H = .65625/3 = .22$

A customary criterion is that individual items should have H values $> .30$. Strong total scale values are $> .50$ and moderate ones $> .40$. In this case, either item 2 (B) or 3 (C) creates trouble and the scale as constituted is not stable.

ProGamma (Now Science Plus Group) has a software catalog that provides Mokken Scale Analysis called MSP for Windows (see “Using the Internet,” p. 208). A good example of the application of Mokken Scaling to psychiatric diagnosis can be found in de Jong and Molenaar (1987).

Dominance Theory of Order

This theory is used with dichotomously scored items that suggest the dominance of one choice over the other (right > wrong or before > after), scored **1** or **0**. The idea is that if someone can answer item 1 correctly and not item 2 then item one should come before item 2. The logical pattern of the number of correctly answered questions versus the number of incorrectly answered can be used in a number of ways. These can consist of determining the sequence of skills necessary for reading, in child development, in designating a hierarchy of cognitive processes, etc. As long as one can develop a key which defines "acceptable" correct answers, the **Order Dominance Model** can be utilized.

The model, based on the early work of Krus, Bart, and Airasian (1975), attempts to build a **chain** or network of items indicating their relative dominance. The steps are as follows:

From a survey or test each question is marked correct or incorrect for each respondent. Every correct answer is marked **1** and incorrect **0**. After ordering the items based on their total scores **all pairs** of questions are compared. If the first question is mostly right and the second question is mostly wrong it is assumed that the first question is a prerequisite for the second and is in the correct order. The researcher looks for pairs of confirmatory responses whose pattern is (1 - 0), and for disconfirmatory patterns (0 - 1). The pairings (0 - 0) and (1 - 1) do not help decide dominance and they are not considered in the initial analysis but are used in Fisher's Exact Test (Finney, 1948).

Suppose five judges respond to six items and their responses are scored and presented as in Table 5.5.

Table 5.5
Response Profiles Over 6 Items

Judge	Items						Σ
	1	2	3	4	5	6	
A	1	0	0	1	1	0	3
B	1	0	1	0	0	0	2
C	1	1	1	1	0	0	4
D	1	0	0	1	0	1	3
E	0	1	1	0	0	0	2
Σ	4	2	3	3	1	1	14

In large samples it is useful to eliminate response pattern profiles which constitute only a small proportion of the data, say less than 5 or 10%, before proceeding farther.

Next, as in Guttman Scaling, the rows and columns are rearranged based on the magnitude of their sums.

Table 5.6
Rearranged Profiles

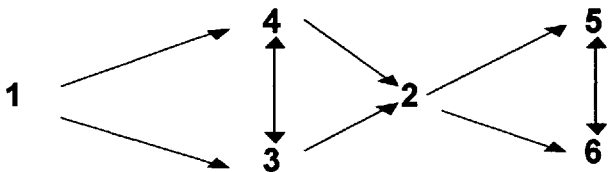
Judge	Items						Sum
	1	3	4	2	5	6	
C	1	1	1	1	0	0	4
*A	1	0	1	0	1	0	3
D	1	0	1	0	0	1	3
B	1	1	0	0	0	0	2
E	0	1	0	1	0	0	2
Sum	4	3	3	2	1	1	14

The six items are paired in all possible (15) ways and each pair is examined to determine if the pairing is confirmatory or disconfirmatory over the five judges. As shown in Table 5.6, Subjects *A's responses to items 1 and 3 are: item 1 correct, Score = 1 and item 3 incorrect, Score = 0. The order is confirmatory. Table 5.7 details the frequencies of the 1-0 and 0-1 responses and sums their differences. The 1-1 and 0-0 frequencies aid in calculating Fisher's exact test.

Table 5.7
Confirmatory (1-0) and Disconfirmatory (0-1) Frequency by Pairs

Items	1-3	1-4	1-2	1-5	1-6	3-4	3-2	3-5	3-6	4-2	4-5	4-6	2-5	2-6	5-6
C (1-0)	2	1	3	3	3	2	1	3	3	2	2	2	2	2	1
D (0-1)	1	0	1	0	0	2	0	1	1	1	0	0	1	1	1
C—D	1	*1	2	3	3	0	1	2	2	1	2	0	1	1	0
(1-1)	2	3	1	1	1	1	2	0	0	1	1	1	0	0	0
(0-0)	0	1	0	1	1	0	2	1	1	1	2	2	2	2	3
Fisher's One-tail Probabilities															
Fisher's p	.60	.40	.40	.80	.80	.30	.30	.40	.40	.70	.30	.60	.60	.60	.80

A preliminary network (chain) can be made in which zero (0) differences assume the items are unrelated, that is, no logical dominance



The frequencies for confirmatory and disconfirmatory responses are placed in an item by item matrix in which the confirmatory responses are placed in the upper triangle and disconfirmatory in the lower triangle as shown in Table 5.8.

Table 5.8
Frequencies of Response

		Confirmatory					
		1	3	4	2	5	6
Disconfirmatory	Items						
	1		2	1	3	3	3
	3	1		2	1	3	3
	4	0	2		2	2	2
	2	1	0	1		2	2
	5	0	1	0	1		1
	6	0	1	0	1	1	

In this case there are 42 total dominance responses, 32 confirmatory and 10 disconfirmatory. The frequencies are converted to proportions by dividing by the number of judges (5) and their sums and differences determined. These are presented in Table 5.9. For items 1 and 3, for example: $d_{13} - d_{31} = 2/5 - 1/5 = .20$ and $d_{13} + d_{31} = 2/5 + 1/5$ or .60.

Table 5.9
Sums and Differences in Proportions of Response

		Confirmatory – Disconfirmatory Proportions $d_{ij} - d_{ji}$					
		1	3	4	2	5	6
Confirmatory +Disconfirmatory Proportions $d_{ij} + d_{ji}$	Items						
	1		.20	.20	.40	.60	.60
	3	.60		.00	.20	.40	.40
	4	.20	.80		.20	.40	.40
	2	.80	.20	.60		.20	.20
	5	.60	.80	.40	.60		.00
	6	.60	.80	.40	.60	.40	

It is useful to provide a significance test on the items to aid in the decision of accepting or rejecting the item relationships. It is possible that some answers were the result of guessing, fatigue, or misreading. One possible measure of statistical significance is a Z score. The ratio of the difference to the square root of the sum is used in McNemar's critical ratio for the calculation of Z:

$$Z_{ij} = (d_{ij} - d_{ji}) / (d_{ij} + d_{ji})^{1/2}$$

$$\text{for } Z_{13} = .20 / (.60)^{1/2} = .26.$$

In Table 5.10 below are the Z values for the order dominance matrix. This model gives the researcher the power to decide how many disconfirmatory responses he or she is willing to keep when building a prerequisite list. The larger the Z score, selected as a cutoff, the more disconfirmatory responses will be included. This cutoff decision is made subjectively. A Z of over 2.5 means the researcher would like to include 95% of all disconfirmatory pairings. The higher the Z score one selects the more likely recursiveness will occur in the dominance lists and the prerequisite list will have fewer levels.

Table 5.10

Z values for the Order Dominance Matrix

Items	1	3	4	2	5	6
1		.26	.45	.63	.77	.77
3			.00	.45	.45	.45
4				.26	.63	.45
2					.26	.26
5						.00
6						

A listing can be made for the items using the Z values as estimates of each item's dominance:

Item 1 > 3, 4, 2, 5, 6

Item 3 = 4

Item 3 > 2, 5, 6

Item 4 > 2, 5, 6

Item 2 > 5, 6

Item 5 = 6.

A hierarchical graph or chain can be formed from the data. In this chain item 1 is the easiest and seen as prerequisite for all other items. Items 3 and 4 and 5 and 6 are independent of each other.

CD-ROM Example Using ORDER

The **ORDER** analysis program (**order.exe**) analyzes the data with three different models. Model C, which is based on using Z scores is applied to the data of Table 5.5. A very small Z score (.25) and a moderate Z score (1.58) are used for comparison purposes.

order2.cfg (see readme file)

Order Program with Data from Table 5.5
 5 6 2 2 .25 1.58
 order2.dat
 order2.out

order2.dat

1 0 0 1 1 0
 1 0 1 0 0 0
 1 1 1 1 0 0
 1 0 0 1 0 1
 0 1 1 0 0 0

order2.out

Order Program with Data from Table 5.5

Number of subjects: 5
 Number of items: 6
 z_values: 0.25 1.58

Model C

Model C (Z-Value = 0.25)

Item Prerequisites

 1 ==> 3 4 2 5 6
 3 ==> 2 5 6
 4 ==> 2 5 6
 2 ==> 5 6
 5 ==>
 6 ==>

Model C (Z-Value = 1.58)

Item Prerequisites

 1 ==> 2 5
 3 ==>
 4 ==>
 2 ==>
 5 ==>
 6 ==>

One can see that if the cut off values are small the prerequisites are well delineated but are more subject to error. With larger values of Z the prerequisites list shrinks but has greater probability of being correct.

Fisher’s Exact Probability

The degree of recursiveness of the 0-1 and 1-0 responses can be studied using Fisher’s exact probability test (Finney, 1948). This includes the 1-1 and 0-0 responses as well. Extremely easy or very difficult items will result in a deflated number of (0-1) responses. This is because the likelihood is that $A > B$ increases as A and B are different in item difficulty. Very easy or very difficult items would result in a smaller number of counter dominances, that is, (0-1). Fisher’s exact probability can provide an indication of the degree of recursiveness or circularity in the data. When the (1-0) and (0-1) responses dominate in the four fold table and are evenly distributed in frequency, the probability of their occurrence is the smallest. For Example using one-tail probabilities:

Item	j				Item	j				Item	j				Item	3			
		1	0				1	0				1	0				1	0	
i	1	a	b		i	1		3		i	1	3	1		2	1	1	2	
	0	c	d			0		2			0	1				0	2	1	
Model					Recursive p = .10					Less recursive p = .80					Example p = .30				

Where $p = (a + b)! (c + d)! (a + c)! (b + d)! / a! b! c! d! N!$

In Table 5.11, Fisher’s probabilities for the items in the example are presented. One can observe that lower probabilities are associated with items with larger frequencies of (1-0) and (0-1) responses. Item 3, for example has an average $p = .52$. The number of agreements, that is, the diagonal values (1-1) and (0-0) for item 3 compared with other items is small.

Table 5.11
Fisher Probabilities

Items	1	4	3	2	5	6	Avg.
1		.80	.60	.40	.80	.80	.68
4	.80		.30	.70	.30	.60	.54
3	.60	.30		.30	.40	.40	.52
2	.40	.70	.30		.60	.60	.52
5	.80	.30	.40	.60		.80	.58
6	.80	.60	.40	.60	.80		.64

$$\begin{aligned}
 v &= 33 \\
 v_c &= 38.2 \\
 v_m &= 21 \\
 CT3 &= .3023
 \end{aligned}$$

Rescaling Reliability

One can rescale the CT3 index as well as KR-20 or KR-21 by adding a unit of (1) to the value found and dividing by two (2). This essentially eliminates negative reliabilities for such indices.

$$\begin{aligned}
 CT3 \text{ rescaled} &= \\
 (CT3 + 1.0)/2.0 &= \\
 (.3023 + 1.0)/2.0 &= .651
 \end{aligned}$$

CT3 values should normally range between -1 and 1, although outliers beyond these limits are mathematically possible. Swartz (2002) enumerated CT3 distributions for 5 objects by 5 judges. She then compared CT3 to KR-21 reliability indices in this range. As shown in Table 5.13, scaled CT3 = .90 while scaled KR-21 = .66 for the alpha = .05 level. When scaled CT3 reaches .99, then scaled KR-21 = .78, alpha = .01. CT3 may be a more appropriate measure of consistency for some kinds of binary data.

Table 5.13
Critical Values for CT3 versus KR-21 for 5 Objects and 5
Judges at the 80-99 Percentile Levels

Cumulative %	CT3 (-1 to 1)	CT3 (0 to 1)	KR21 (0 to 1)
80	.02	.60	.50
90	.36	.79	.57
95	.50	.90	.66
99	.84	.99	.78

Application Example

An example serves to illustrate how CT3 can give a different view of consistency than KR-21. Table 5.14 contains data from two sets of sixth grade students who participated in a summer camp for children exploring space and time concepts (Gelphman, Knezek, & Horn, 1992). This is a robotic imaging technology in which the image-capture view and the path of the camera can be modified to yield perspectives different from the normal human eye. Two sets of students responded to five questions about space and time. As shown in Table 5.14, both sets of students had high reliability indices for these five questions (CT3 = 1.0). KR-21, however, was .80 and .85 for the two sets, respectively. No item-inconsistent response patterns were displayed in either set of data, so it seems reasonable that the consistency index should be "1.0." CT3 produces the more intuitively valid indicator in this case. The **ORDER** program on the CD-ROM calculates CT3.

Table 5.14
Right/Wrong Data for Ten Students Completing
Topocam-Enriched Content Lesson

Obs		COL1	COL2	COL3	COL4	COL5	ROWT
11	Group 1	1	1	1	1	1	5
12	Group 1	1	1	0	1	1	4
13	Group 1	0	1	0	1	1	3
14	Group 1	0	0	0	0	1	1
15	Group 1	0	0	0	0	0	0
16	Group 2	1	1	1	1	1	5
17	Group 2	1	1	1	1	1	5
18	Group 2	1	1	0	1	1	4
19	Group 2	0	1	0	1	1	3
20	Group 2	0	0	0	0	0	0

FINAL -KR VALIDATED &
CT3 VALIDATED

	CT3	KR21
Group 1	1.00	0.80
Group 2	1.00	0.85

Partial Correlations as a Measure of Transitivity

Douglas R. White (1998) has written a computer program for Multidimensional Guttman Scaling. In viewing dominance, White uses the term entailment between A and B. By this he means B is easier than A, B is a subset of A, if A then B, or A entails B. Using his program, as many as 50 dichotomous items can be analyzed to determine a network of transitive entailments.

His program is built on a test of the assumption of transitivity. If A exceeds B, and B exceeds C then we infer A exceeds C. A exceeds C if A, B, C pass the tests of transitivity. The sum of the exceptions to such transitivity is called the strong measure of transitivity. For example, if a 1 precedes a 0 in a set whose sums are ordered from small to large, this is an exception to transitivity. In the table below exceptions occur in the fifth, seventh, and ninth rows.

Data				
	A	B	C	D
	1	1	1	1
	1	1	1	1
S	1	1	1	1
U	0	1	1	1
B	1	0	1	1
J	0	0	1	1
E	0	1	0	1
C	0	0	0	1
T	0	0	1	0
S	0	0	0	0
	0	0	0	0
	0	0	0	0
a	4	5	7	8

For a weak measure of transitivity White uses the partial correlation, finding, for example the partial r between A and B controlling for C: $r_{AB \cdot C} = r_{AB} - r_{AC}r_{BC}/[1 - r_{AC}^2)(1 - r_{BC}^2)]^{1/2}$. Zero or positive partial correlations are indications of transitivity. White's analysis is based on using a two by two table. If one looks at items A and B and counts the number of times the responses are 1-1, 1-0, 0-1 and 0-0, these are entered as:

		B	
		1	0
A	1	e	d
	0	f	g

		B	
		1	0
A	1	3	1
	0	2	4

The correlation between A and B is .408. A implies B exceptions are cell d or 1 and the percent is d/N or 1/10 or .10. B implies A exceptions are cell f or 2 and the percent is .20. If A implies B is less than or equal to B implies A, it is defined as a Strong inclusion. Using the column sums of the initial data (4, 5 ,7, 8) random distributions are created and compared to the original, using signal detection to distinguish entailments from statistical noise. Finally, a set of suggested chains is given. One can experiment with this program which can be downloaded free at:

<http://eclectic.ss.uci.edu/~drwhite/entail/manual.html>.

6

COMPARATIVE JUDGMENT

Thurstone (1927) provided a rationale for ordering objects on a psychological continuum. Psychological objects are stimuli for which some reaction takes place within the sensory system of the individual. The objects can be a beautiful girl, a telephone's ring, sandpaper, sugar water, or nitrous oxide. They may also include statements, such as "I hate school," or "I am a patriot".

Attitudes are Normally Distributed

Thurstone postulated that for any psychological object: (1) reactions to such stimuli were subjective; and (2) judgment or preference for an object may vary from one instance to another. Thurstone suggested that, although we may have more or less favorable reactions to a particular psychological object, there is a most frequent, or typical, reaction to any object or stimulus. The most frequent reaction is called the modal reaction. The mode can be based on repeated reactions of a single individual or the frequency of the reactions of many subjects.

Because the normal curve is symmetrical, the most frequent reaction (the mode) occupies the same scale position as the mean. Thus, the mean can also represent the scale value for the particular psychological object. In his simplest case, Case V, Thurstone assumed that reactions to various stimuli were normally distributed. He also assumed that the variance of the reactions around each mean would be the same. Scale values can be acquired, however, only within a relative frame. It is necessary, therefore, to have at least two objects so that a comparison can be made. Figure 6.1 illustrates this case.

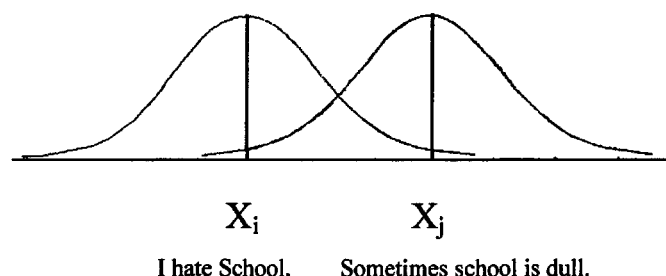
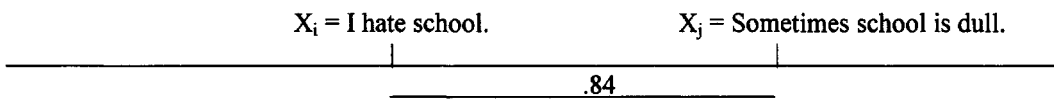


FIG. 6.1. Theoretical normal distribution about two different psychological objects.

Suppose X_i and X_j are two psychological objects that are to be judged on a continuum of positive affect toward school. Suppose X_i is "I hate school," and X_j is "Sometimes school is dull." We might ask a group of participants to judge which statement is more favorable toward school attendance. If 80% of the subjects choose j as more favorable than i and therefore only 20% choose i as more favorable than j , we might argue that the average reaction to j should be higher on a scale than the average reaction to i or $X_j > X_i$. The separation between X_j and X_i is a function of the number of times j is rated over i . Using paired comparisons, we can count the votes and get proportions of preference. If, with 50 subjects, X_j (School is dull.) is chosen 40 times over X_i (I hate school.), then the proportion is 40/50 or .80.

Thurstone's Case V

Proportions can be expressed as normal variates (i.e., Z standard scores can be obtained for proportions). In this case the normal deviate $Z_{ij} = .84$ (for $p = .80$). The scale separation between two psychological objects can be made in terms of this normal variate. Diagrammatically, we can say that somewhere on the continuum of attitude toward school (attendance) X_j and X_i are separated by a distance of .84 as follows:



Note that the mean of the distribution of responses around the stimuli will never be known. The difference between any two means, however, can be estimated if one makes the assumption of normality mentioned previously. Thurstone's use of the normal variate as a measure of distance (Case V) is justified in the following way:

Following Thrustone (1927), a test of the difference between means of two normal distributions is:

$$Z_{12} = \bar{X}_1 - \bar{X}_2 / \sqrt{(S_1^2 + S_2^2 - 2rS_1S_2)}$$

Where S represents the standard error and r is the correlation between the two groups. Modern researchers would suggests that this is the formula for Student's t .

Thurstone solves this equation for the difference between the means and then letting the means represent the scale values of two stimuli (the mean and the mode are the same in a normal distribution) and by assuming the items to be uncorrelated (i.e., $r = 0$), the formula reduces to:

$$\bar{X}_1 - \bar{X}_2 = Z_{12} \sqrt{(S_1^2 + S_1^2)}$$

By assuming that the variances of response are equal for the two items, the value under the radical becomes a constant and assuming the constant to be one (1) in this case (Thurstone's Case V) the formula reduces to:

$$\bar{X}_1 - \bar{X}_2 = Z_{12}$$

Case V Example

Thurstone's procedure for finding scale separations starts with the votes derived from some paired comparison schedule of objects. The votes can be accumulated in a square array by placing a 1 in each row and column intersection in which the column object is judged or preferred over the row object.

A matrix can accumulate a large number of persons' responses to the objects. In the fictitious example that follows, the first table (Table 6.1) contains the frequency of choices of 100 student's to the psychological objects **cafeteria**, **gymnasium**, **theater**, **library**, **classroom**. The students were asked to judge the importance of each to their college education. The objects were paired in the 10 possible ways and votes accumulated in a frequency matrix.

Table 6.1
Accumulated Frequency (F_{IJ} Matrix), $N = 100$

	class	cafe	gym	lib	the
classroom		20	30	35	10
cafeteria	80		30	40	20
gymnasium	70	70		45	15
library	65	60	55		25
theater	90	80	85	75	
Sum	305	230	200	195	70

Note: Each entry contains the votes of column objects over the row objects

Initially, the column sums are found and if the sums are not in order (as shown) the rows and columns of the matrix are rearranged so that the column sums are ordered from **smallest** to **largest**. Under the variance stable or simplified rank method we would proceed directly to use the sum of the votes as scale scores. But, under Thurstone's rationale, the individual frequencies are converted to proportions as shown in Table 6.2

Table 6.2
Proportions F_{IJ}/N

	the	lib	gym	cafe	class
the	.50	.75	.85	.80	.90
lib	.25	.50	.55	.60	.65
gym	.15	.45	.50	.70	.70
cafe	.20	.40	.30	.50	.80
class	.10	.35	.30	.20	.50

A proportion of .50 is placed on the diagonal of this matrix under the assumption that any object judged against itself would receive a random number of votes. The expectation is that 50% of the time, the subject would choose the column object and 50% of the time the row object. Note that the objects have been rearranged according to the sums in Table 6.1. Next, the

proportions are converted to normal deviates by reference to the normal distribution as shown in Table 6.3. Normal values are readily available on the Internet.

Table 6.3

Normal Deviates

	the	lib	gym	cafe	class
the	.00	.67	1.03	.84	1.28
lib	-.67	.00	.13	.25	.38
gym	-1.03	-.13	.00	.52	.52
cafe	-.84	-.25	-.52	.00	.84
class	-1.28	-.38	-.52	-.84	.00

*Normal deviates are found in any statistical text and on the internet or in any statistics text.
(For PQRS use www.eco.rug.nl/medewerk/knypstra/)

Finally, the differences between column stimuli are found as shown in Table 6.4. If the data are complete, the differences between the column sums of the normal deviates are equal to the sums of the column differences.

Table 6.4

Column Differences

	lib - the	gym - lib	cafe - gym	class - cafe
	.67	.37	-.19	.44
	.67	.13	.12	.13
	.90	.13	.52	.0
	.59	-.27	.52	.84
	.90	-.14	-.32	.84
Sum	3.73	.22	.65	2.25
K	5	5	5	5
Average	.746	.044	.13	.45

Knowing the differences among the objects we can assign scale values to each by accumulating the differences or distances among them. Therefore:

theater	the	= 000	= .00
library	lib	= 000 + .75	= .75
gymnasium	gym	= 000 + .75 + .04	= .79
cafeteria	cafe	= 000 + .75 + .04 = .13	= .92
classroom	class	= 000 + .75 = .04 = .13 + .45	= 1.32

A graphical representation can be made as follows:

theater	lib	gym	cafe	class
0			1.00	2.00

Should proportions greater than .98 occur in the data, they are reduced to .98. This is similarly true for proportions less than .02, which are made equal to .02. The reason for this restriction is that normal deviations for extreme proportions usually result in an extreme distortion of the scale values (100% equals infinity, for example). If data are missing, the entries are left blank and no column differences are found for the blank entries. Averages of column differences are then found by dividing by k (the number of stimuli) reduced by the number of incomplete entries. The Case V method requires assumptions of equal dispersion of reactions and a lack of correlation between judgments of different objects. If these assumptions cannot be met, some other method or case may have to be used. The Case V is the simplest and most frequently used of the various cases that Thurstone explored.

Reliability

A test of the effectiveness of any linear scale can be based on the ability of the scale scores to recapture the original proportions or frequencies used to produce the scale. Traditionally (for the Case V model) this is done by converting the Z scale values to obtained proportions (p'). That is, finding differences between all pairs of Z scale values and converting each difference into a proportion. Then the average difference between the original (p) and obtained proportions (p') is calculated. A measure, called the average deviation (AD) of the two sets of proportions, is then used as the index of scalability. Smith (1968), did a Monte Carlo study of the distribution of the AD and confirmed its relationship to Mosteller's (1951) X^2 test.

Gulliksen and Tukey (1958) have approached the problem of scale reliability using analysis of variance because they wished to answer the question of how effectively the scale scores account for variability in responses (i.e., what percentage of the total variance is accounted for). In this case the traditional definition of reliability, $r_{tt} = 1 - S_e^2/S_T^2$ (where S_e^2 = error variance and S_T^2 = total variance) is used as an index of scalability.

Application: Seriousness of Crimes: Then and Now

Mahoney (1986) had 62 college students judge the seriousness of the same 19 crimes first reported on by Thurstone (1927), Thurstone's study was replicated by Coombs (1964). The crimes, **abortion, adultery arson, assault, battery, bootlegging, burglary, counterfeiting, embezzlement, forgery homicide, kidnapping, larceny, libel, perjury, and rape** were paired in all possible (171) ways and presented according to Ross (1934). The resulting data were processed using Gulliksen's (1958) COMPPC program which analyzed the data under the Case V model and provided a reliability measure (Gulliksen and Tukey, 1958). For the 1986 scale, the reliability was .88. The results are shown in Fig. 6.2. Abortion, seduction, smuggling and adultery have continued their downward trend in perceived seriousness

Case V Program

Thurstone's method is not often used, but its historical perspective makes later methods more understandable. The Case V method can be found in the PC-MDS programs (called Case5) promulgated by the Department of Marketing at Brigham Young University over the internet (Webpage = <http://marketing.byu.edu>). The FORTRAN listing for Complete Paired Comparisons can be found in Scaling Methods (Dunn-Rankin, 1983).

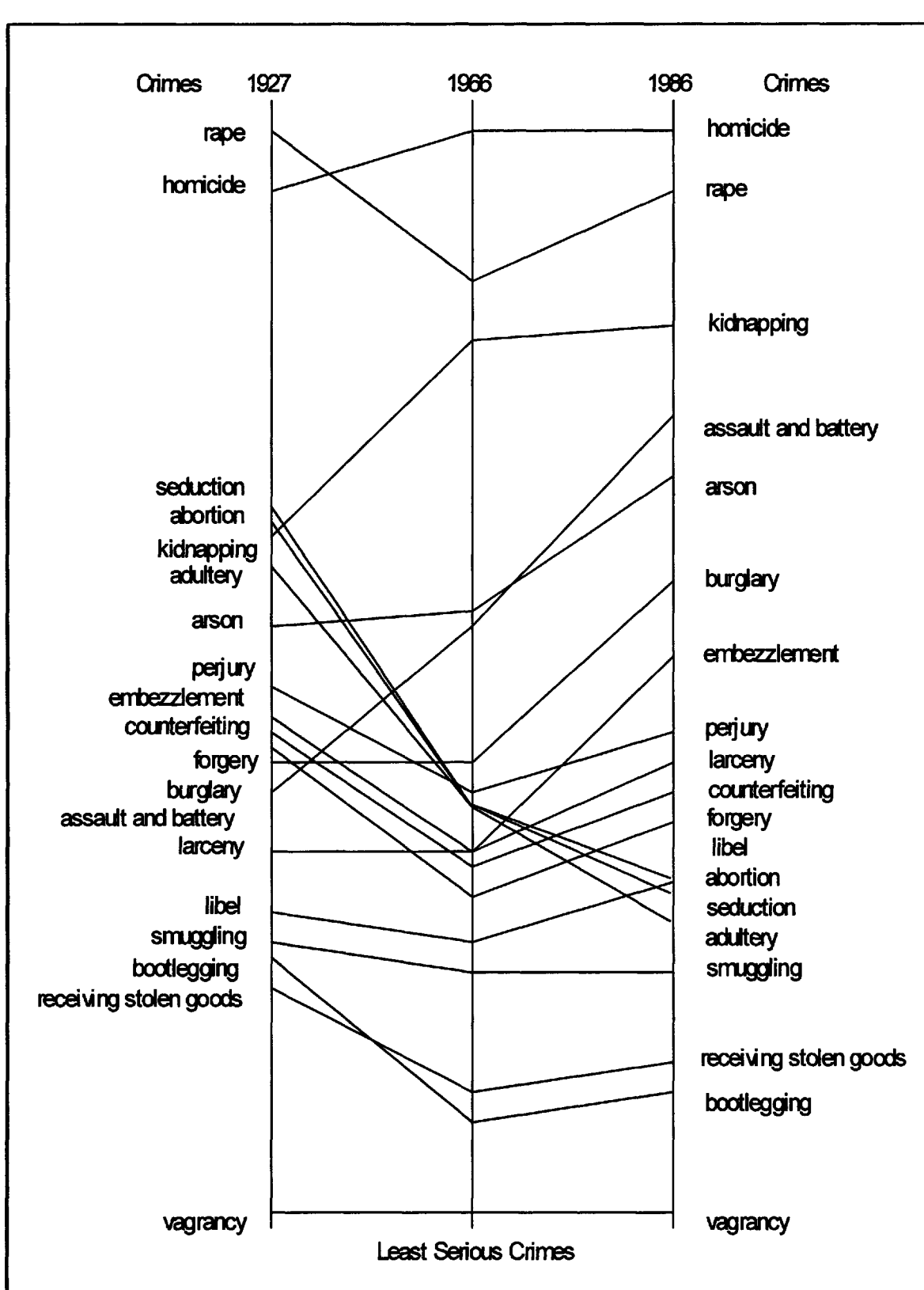


FIG. 6.2. College student perception of seriousness of crimes 1927, 1966, and 1986.

7

CATEGORICAL RATINGS

The most popular unidimensional method of attitude measurement involves ordered categories. More traditional names such as **Summated Ratings**, **Likert Scaling** and **Semantic Differential** are forms of ordered categories. In this method, the judges are asked to place items in a fixed number of categories, usually 2, 3, 4, 5, 7, 9, or 11. A typical example of this format is given in Table 7.1. A unidimensional scale of attitude toward reading is proposed for these eight statements. Judges are asked to indicate the degree of positive affect toward reading for each statement by marking appropriately. It is clear that the format can accommodate a great many statements and it calls for only one action per statement by each judge. It is the accumulation of the responses of a number of judges that provides the data for creating the scale.

Table 7.1

Example of an Ordered Category Rating Scale

	No						Yes
	1	2	3	4	5	6	7
1. I try reading anything I can get my hands on.	_____	_____	_____	_____	_____	_____	_____
2. I read when there is nothing else to do.	_____	_____	_____	_____	_____	_____	_____
3. When I become interested in something I read about it	_____	_____	_____	_____	_____	_____	_____
4. I have never read an entire book.	_____	_____	_____	_____	_____	_____	_____
5. I seldom read anything.	_____	_____	_____	_____	_____	_____	_____
6. I almost always have something I can read.	_____	_____	_____	_____	_____	_____	_____
7. I never read unless I have to.	_____	_____	_____	_____	_____	_____	_____
8. I only read things that are easy.	_____	_____	_____	_____	_____	_____	_____

Because the formulation of survey instruments is easily made, abuses of the ordered category method are frequently found. Some of the more common abuses are: **First**, it is rare that judgments are initially sought by the experimenter. Rather the category headings ask for degrees of personal agreement. If the use of such an instrument is to be valid and reliable one must speculate that agreement and judgment are similar and that the trial sample is similar to the final population. These two speculations are not always justified. **Second**, the statements formulated are seldom unidimensional in character yet are analyzed as if they were. **Third**, an assumption of equality of intervals is made. For example, the value of the distance between 5 and 4 and the distance between 4 and 3 (in Table 7.1) is assumed to be equal and is usually assigned a value of 1 (see p. 21, naming categories).

Green's Successive Categories

The scaling method of successive intervals (Green, 1954) is an attempt to accommodate more items than other unidimensional techniques can and to estimate the distance or interval between the ordered categories. When a number of judges have marked the items, a distribution of judgments for each item is created (see, for example, Table 7.2). In this method, the average of the normal deviates assigned to the cumulative proportions of responses in each category represents the scale score of the item **but only after each deviate is subtracted from the category boundary**. As in the Case V model (see previous chapter), variances around scale values are assumed to be equal.

Table 7.2
Frequencies of Response by 15 Judges to Reading Attitude Statements

Statements	No						Yes
	1	2	3	4	5	6	7
1	0	0	1	3	3	1	7
2	2	1	3	4	0	1	4
3	0	0	0	1	1	5	8
4	8	4	1	2	0	0	0
5	10	4	0	0	1	0	0
6	0	2	0	1	1	4	7
7	8	4	1	0	1	1	0
8	4	5	0	4	1	0	1

The boundaries of the intervals are located under the assumption that the judgments for each item are distributed normally. In order to analyze the items under the cumulative normal distribution, the categories are arranged from **least** to **most** favorable and the cumulative frequency distributions are found, as in Table 7.3.

Table 7.3
Cumulative Frequency Distributions

Statements	No						Yes
	1	2	3	4	5	6	7
1	0	0	1	4	7	8	15
2	2	3	6	10	10	11	15
3	0	0	0	1	2	7	15
4	8	12	13	15	15	15	15
5	10	14	14	14	15	15	15
6	0	2	2	3	4	8	15
7	8	12	13	13	14	15	15
8	4	9	9	13	14	14	15

These summed frequencies are then converted to cumulative proportions, as shown in Table 7.4.

Table 7.4
Cumulative Proportions

Statements	Categories						Yes
	No						
	1	2	3	4	5	6	7
1			.07	.27	.47	.53	1.00
2	.13	.20	.40	.67	.67	.73	1.00
3				.07	.13	.47	1.00
4	.53	.80	.87	1.00	1.00	1.00	1.00
5	.67	.93	.93	.93	1.00	1.00	1.00
6		.13	.13	.20	.27	.53	1.00
7	.53	.80	.87	.87	.93	1.00	1.00
8	.27	.60	.60	.87	.93	.93	1.00

Any proportions greater than .98 or less than .02 are rejected and the cumulative proportions are converted into normal deviates by referring to areas of normal distributions (see Table 7.5). The cumulative proportions are converted to Z scores

Table 7.5
Unit Normal Deviates (Z scores)

Statements	Categories						Yes
	No						
	1	2	3	4	5	6	7
1			-1.48	-.61	-.08	.08	
2	-1.13	-.84	-.25	.44	.44	.61	
3				-1.47	-1.13	-.08	
4	.08	.84	1.13				
5	.44	1.48	1.48	1.48			
6		-1.13	-1.13	-.84	-.61	.08	
7	.08	.84	1.13	1.13	1.48		
8	-.61	.26	.26	1.13	1.48	1.48	

The differences between the categories for each item are found and the average of the differences is equal to the boundary between the two columns. For missing entries **no differences** are found and the average is found for those items for which a difference is available. See Table 7.6. The averages in the bottom row of Table 7.6 are the distances or intervals between the categories.

Table 7.6
Matrix of Differences

Statements	2-1	3-2	4-3	5-4	6-5
1			.87	.53	.16
2	.29	.59	.69	.00	.17
3				.34	1.05
4	.76	.29			
5	1.04	.00	.00		
6		.00	.29	.23	.69
7	.76	.29	.00	.35	
8	.87	.00	.87	.35	.00
Sum	3.72	1.17	2.72	1.80	2.07
n	5	6	6	6	5
Average	.74	.20	.45	.30	.41

By setting the first boundary arbitrarily as $B_1 = 0$, the remaining boundaries can be computed by summing the boundaries cumulatively from left to right as follows:

$$B_1 = 0$$
$$B_2 = 0 + .74$$
$$B_3 = 0 + .74 + .20$$
$$B_4 = 0 + .74 + .20 + .45$$
$$B_5 = 0 + .74 + .20 + .45 + .30$$
$$B_6 = 0 + .74 + .20 + .45 + .30 + .41$$

$$= .00$$
$$= .74$$
$$= .94$$
$$= 1.39$$
$$= 1.69$$
$$= 2.10$$

A comparison between an equal interval assumption and the boundaries obtained is shown by dividing the sum (2.10) by five this yields equal intervals of .42.

Equal Intervals	1	2	3	4	5	6
Actual Intervals	1	2	3	4	5	6

This illustrates that the length between 1 and 2 is quite long whereas the interval between 2 and 3 is fairly small. In order to obtain the scale scores, the normal deviate values (given in Table 7.5) are subtracted from the category boundaries. In this case the first boundary is taken as zero and no values are found for column 7. Table 7.7 shows this calculation. The row sums are then averaged.

Table 7.7
Boundaries Minus Column Normal Deviates

Statements							Scale		
	$B_1 - 1$	$B_2 - 2$	$B_3 - 3$	$B_4 - 4$	$B_5 - 5$	$B_6 - 6$	Sum	k	Average
1			2.42	2.00	1.77	2.02	8.21	4	2.05
2	1.13	1.58	1.19	.45	1.25	1.49	7.59	6	1.27
3				2.86	2.82	2.18	7.86	3	2.62
4	-.08	-.10	-.19				-.37	3	-.12
5	-.44	-.79	-.54	-.09			-1.81	4	-.45
6		1.87	2.07	2.23	2.30	2.02	10.49	5	2.10
7	-.08	-.10	-.19	.26	.21		.10	5	.02
8	.61	.48	.68	.26	.21	.62	2.86	6	.48

These scale values or scores indicate that the reading items should be arranged as follows:

- 3. When I become interested in something I read a book.
- 6. I almost always have something I can read.
- *1. I try reading anything I can get my hands on.
- 2. I read when there is nothing else to do.
- 8. I only read things that are easy.
- 7. I never read unless I have to.
- *4. I have never read an entire book.
- 5. I seldom read anything.

Starred items may be deleted from final scale. The experimenter may wish to delete items close together on the scale. In this study, the experimenter might eliminate items 1 and 4 from the final scale.

Discussion

The Program **TSCALE** formulated by Veldman (1967) performs Successive Interval Scaling and is provided on the CD-ROM. Veldman's program differs slightly from Green's method but the relative positions of the objects remain much the same. When using **TSCALE**, if **complete** endorsement occurs for an item, (for example all the respondents strongly agree with the statement) then the variance of the item is zero (0) and it will appear at the zero point of the scale even though it may have universal endorsement and should have the highest scale value or the lowest. Such items cannot be scaled but may provide insight into the upper or lower boundaries of the scale. The FORTRAN program for **TSCALE** can be found in *Scaling Methods* (Dunn-Rankin, 1983). If an ordered category instrument has a large number of items, it may be expected that the instrument contains more than one unidimensional scale. If this is expected, the multidimensional methods of clustering, factor analysis, or multidimensional scaling methodologies may be used (See PART III and PART IV of this text).

TSCALE Analysis of Reading Attitude

The reading attitude data were submitted to the TSCALE program on the CD-ROM and the results are given below. There are slight differences with Green’s method due to rounded Z values and the location of the zero point (see the relative positions of items 1 and 6).

tscale.cfg

Reading Response Attitude Statements	Title
15 8 7	Parameters (Judges, Items, Categories)
tscale.dat	
tscale.out	

tscale.dat

3 1 4 1 1 2 1 1	}Data
4 1 5 1 1 2 1 1	
4 2 6 1 1 4 1 1	
4 3 6 1 1 5 1 1	
5 3 6 1 1 6 1 2	
5 4 6 1 1 6 1 2	
6 4 7 1 1 6 1 2	
7 4 7 2 1 7 2 2	
7 4 7 2 1 7 2 4	
7 6 7 2 2 7 2 4	
7 7 7 2 2 7 2 4	
7 7 7 3 2 7 3 4	
7 7 7 4 2 7 5 5	
7 7 7 4 5 7 6 7	
5 3 6 1 1 6 1 2	

tscale.out

TSCALE Output
Reading Response Attitude Statements
ITEMS = 8 JUDGES = 15 CATEGORIES = 7

SCALE VALUES				FREQUENCIES							
RANK-ORDERED		ITEM-ORDERED		1	2	3	4	5	6	7	
3	3.13	1	2.48	0	0	1	3	3	1	7	3. When I become interested in something I read a book about it.
1	2.48	2	1.55	2	1	3	4	0	1	4	1. I try reading anything I can get my hands on.
6	2.44	3	3.12	0	0	0	1	1	5	8	6. I almost always have something I can read.
2	1.55	4	-0.06	8	4	1	2	0	0	0	2. I read when there is nothing else to do
8	0.76	5	-0.30	10	4	0	0	1	0	0	8. I only read things that are easy.
7	0.24	6	2.44	0	2	0	1	1	4	7	7. I don't read unless I have to.
4	-0.06	7	0.24	8	4	1	0	1	1	0	4. I have never read an entire book.
5	-0.30	8	0.76	4	5	0	4	1	0	1	5. I seldom read anything.

Summated Ratings

Likert (1932) argued that: (1) category intervals are generally equal; (2) category labels can be preset subjectively and (3) the judgment phase of creating a scale can be replaced by an item analysis performed on the results of the respondents responses. These three arguments mean that in Likert scaling, the strength of a person's preference about all the psychological objects replaces the direction and intensity of the specific objects that a respondent would have judged. **Successive Interval Scaling** and **Likert Scaling**, when carefully applied, often yield similar results. Because Likert Scaling is easier, it is more popular.

An Example of Likert Scaling

The objects are chosen and unit values are assigned to each ordered category, for example, the integers 1 through 7. After subjects respond by checking or marking one of the categories for each item, an N by K (subject by item) matrix of information is generated as shown in Table 7.8. If this was a final instrument, the **Total** attitude scores in the right hand column would be utilized directly. In this instrument, each item correlates positively with the total score.

Table 7.8
Responses to Reading Attitude Survey

Students	Items								Total
	I-1	I-2	I-3	I-4	I-5	I-6	I-7	I-8	
1	3	1	4	1	1	2	1	1	14
2	4	1	5	1	1	2	1	1	16
3	4	2	6	1	1	4	1	1	20
4	4	3	6	1	1	5	1	1	22
5	5	3	6	1	1	6	1	2	25
6	5	4	6	1	1	6	1	2	26
7	6	4	7	1	1	6	1	2	28
8	7	4	7	2	1	7	2	2	32
9	7	4	7	2	1	7	2	4	34
10	7	6	7	2	2	7	2	4	37
11	7	7	7	2	2	7	2	4	38
12	7	7	7	3	2	7	3	4	40
13	7	7	7	4	2	7	5	5	46
14	5	3	6	1	1	6	1	2	25
15	7	7	7	4	5	7	6	7	50
Sum	85	63	95	27	23	86	30	42	453
Mean	5.6	4.2	6.3	1.8	1.53	5.73	2.0	2.8	30.2
S	1.44	2.14	0.90	1.08	2.06	1.75	1.55	1.78	10.6
r_{tot}	.89	.94	.81	.90	.74	.80	.85	.94	

A review of the Attitude Survey data is made through item analysis. Each respondent's categorical value is provided in the body of Table 7.8. Using **SAS PROC Corr Alpha**, the mean (item difficulty) and standard deviation of each item are calculated and Pearson (r) correlations of items with the "total score on all items" are found. The correlation acts as a discrimination index for each item. That is, if the item correlates highly with the total score it is internally consistent. From the item analysis, Item 5 (I seldom read anything) has the lowest correlation (.74) with the total test score. Item 3 has the lowest variability ($S = .89$). Items with low variability may fail to discriminate. Finally, Cronbach's alpha reliability coefficient (Cronbach, 1951) is determined. Despite using only 8 items, Alpha is 0.95 or very consistent. Items are eliminated on the basis of poor internal consistency, very high or low endorsement, or lack of variability. If the items are ordered, based on their sums, the order is:

$$3 > 1 > 6 > 2 > 8 > 7 > 4 > 5$$

This is the same order obtained in using **TSCALE**, Successive Interval Scaling on the CD-ROM.

Discussion

Initial item selection for ordered category scaling can be aided by using the guidelines prescribed in chapter 1, p. 18. One should be careful to avoid "foldback" analysis in which a selection of discriminating items is used to predict differences in the sample from which the items were originally selected. (See Blumenfeld, (1972); "I am never startled by a fish".) The steps used in creating an ordered category scale are as follows:

1. Decide on the number of dimension(s). (If more than one dimension see multidimensional methods.)
2. Collect objects (observe criteria table, p. 18; make a pilot instrument).
3. Make a semantic description and exclude semantic outliers (see p. 109).
4. Present items to judges and obtain their similarity judgments (could be done by free clustering).
5. Find item statistics (mean [proportion passing], S.D., r of item with total test score).
6. Run on **TSCALE** (see CD-ROM) for successive intervals.
7. Revise scales. In the finished scale the category continuum is changed to one of agreement versus disagreement instead of judgments of similarity.

It is always useful to include negatively worded items in ordered category scale construction. This makes the subjects read the items more carefully. When clustering or scoring such items their responses may need to be reversed (see p. 113).

Example: Remmers's General Scale

Remmers (1963) popularized the use of the ordered category scale and produced general scales, (Fig. 7.1). It is interesting to note that Remmers includes some very extreme statements to ensure that all possible representations are available in the final instrument. Also note that each statement is reasonably short and that there are no compound sentences. When surveys like Figure 7.1 are given, a selection of demographics (age, sex, occupation, education, ethnicity,

etc.) relating to the problem being studied are sought. These may then be correlated with the scores of the respondents.

Please fill in the blanks below. (You may leave the space for your name blank.)

Name _____ Boy Girl (circle one), Date, _____ Grade _____

What occupation would you like to follow? _____

Directions: Following is a list of statements about a school subject Place a plus (+) sign before each statement with which you agree with reference to the subjects listed.

English — Math

E M

- _____ : 1. I am crazy about this subject
 - _____ : 2. The very existence of humanity depends upon this subject.
 - _____ : 3. If I had my way, I would compel everybody to study this subject.
 - _____ : 4. This subject is one of the most useful subjects I know.
 - _____ : 5. I believe this subject is the basic one for all school courses.
 - _____ : 6. This is one subject that all Americans should know.
 - _____ : 7. This subject fascinates me.
 - _____ : 8. The merits of this subject far outweigh the defects.
 - _____ : 9. This subject gives pupils the ability to interpret life's problems.
 - _____ : 11. This subject makes me efficient in school work..
 - _____ : 13. This subject is interesting.
 - _____ : 14. This subject teaches methodical reasoning.
 - _____ : 15. This subject serves the needs of a large number of boys and girls.
 - _____ : 16. All methods used in this subject have been thoroughly tested.
 - _____ : 17. This subject has its merits and fills its purpose quite well.
 - _____ : 18. Every year more students are taking this subject.
 - _____ : 19. This subject aims: mainly at power of execution or application.
 - _____ : 20. This subject is not based on untried theories.
 - _____ : 21. I think this subject is amusing.
 - _____ : 22. This subject has its drawbacks, but I like it.
 - _____ : 23. This subject might be worth while if it were taught right.
 - _____ : 24. This subject doesn't worry me in the least.
 - _____ : 25. My likes and dislikes for this subject balance one another.
 - _____ : 26. This subject is all right, but I would not take any more of it.
 - _____ : 27. No student should be concerned with the way this subject is taught.
 - _____ : 28. To me this subject is more or less boring.
 - _____ : 29. No definite results are evident in this subject.
 - _____ : 30. This subject does not motivate the pupil to do better work.
 - _____ : 31. This subject had numerous limitations and defects.
 - _____ : 32. This subject interferes with developing..
 - _____ : 33. This subject is dull.
 - _____ : 34. This subject seems to be a necessary evil.
 - _____ : 36. The average student gets nothing worth having out of this subject.
 - _____ : 37. All of the material in this subject is very uninteresting.
 - _____ : 38. This subject can't benefit me.
 - _____ : 39. This subject has no place in the modern world.
 - _____ : 40. Nobody likes this subject.
 - _____ : 42. This subject is all "junk".
 - _____ : 43. No sane person would take this subject.
 - _____ : 44. Words can't express my antagonism towards this subject.
 - _____ : 45. This is the worst subject taught in school.
-

Fig. 7.1. A scale for measuring attitude toward school subjects.

Application: Revising A Scale

McClarty (1980) selected **A Foreign Language Attitude Scale for revision**. An inspection of the items suggested several incongruent items. The scale was designed to measure "attitude toward learning a (particular) foreign language" yet it contained items like "I would like to be a (Japanese) teacher," and "Everyone in school should take a foreign language." The original scale is given in Figure 7.2. Two hundred and thirty-one high school students (grades 7-12) taking their first year of high school Japanese responded to the scale. Many of these students were of Japanese ancestry and had attended Japanese language schools in their elementary school years.

Foreign Language Attitude Scale

1 - do not agree at all 2 - agree a little bit 3 - agree quite a bit 4 - agree very much

-
1. I would like studying Japanese.
 2. I would like to learn more than one foreign language.
 3. I like to practice Japanese on my own.
 4. Most people enjoy learning a foreign language.
 5. Everyone in school should take a foreign language.
 6. Japanese is interesting.
 7. It is too bad that so few Americans can speak Japanese.
 8. Anyone who can learn English can learn Japanese.
 9. I would like to travel in a country where Japanese is spoken.
 10. The way Japanese people express themselves is very interesting.
 11. Japanese is an easy language to learn.
 12. I would like to be a Japanese teacher.
 13. I would like to take Japanese again next year.
 14. The Japanese I am learning will be useful to me.
 15. I would like to know Japanese-speaking people of my own age.
 16. Students who live in Japanese-speaking countries are just like me.
 17. I'm glad Japanese is taught in this school.
 18. My parents are pleased that I'm learning Japanese.
 19. I like to hear Japanese people talk.
 20. Japanese is one of my most interesting subjects.
 21. Studying Japanese helps me to understand people of other countries.
 22. I think everyone in school should study a foreign language.
 23. Americans really need to learn a foreign language.
 24. What I learn in Japanese helps me in other subjects.
 25. Learning Japanese takes no more time than learning any other subject.
 26. Sometimes I find that I'm thinking in Japanese.
 27. My friends seem to like taking Japanese.
 28. I'm glad that I have the opportunity to study Japanese.
 29. I use Japanese outside the classroom.
 30. I'm looking forward to reading Japanese books on my own.
 31. I would like to study more Japanese during the next school year.
 32. Japanese is one of the most important subjects in the school curriculum.
-

Fig. 7.2. The original foreign language scale which is to be revised.

Because the attitude scale is relatively complex, it was decided to use the semantic description matching design given by Levy and Guttman (1975) as an initial attempt in obtaining unidimensionality. In this technique, an effort is made to classify statements by their semantic items, it is reasonable to redefine items using this method. First, a model sentence, denoting the

dimensions. Although Guttman recommends using this technique in the construction of trial items, it is reasonable to redefine items using this method. First, a model sentence, denoting the attitude being "measured" by the scale, was created (**I would like to study Japanese**) and each scale sentence was compared to the model.

Table 7.9
Semantic Description of Attitude Statements

	Model						Students	
	I	(would) like	to study	Japanese	(qualifier)	sum*	Mean	SD
1	x	x	x	x		4	2.86	.74
2	x	x	x	f.l.		3	2.78	1.67
3	x	x	x	x	on my own	4—	2.15	.84
4	people	x	x	f.l.		2	2.70	.86
5	everyone	should	speak	f.l.		1	2.96	1.11
6	Implied	interest	-	x		3	3.20	.81
7	Americans	should	speak	x		1	2.77	.99
8	Anyone	can	learn	x		1	2.63	1.03
9	x	x	travel	x		3	3.00	1.03
10	Implied	interest	speech	x		2	3.20	.87
11	x	easy	learn	x		2	2.34	.88
12	x	x	teach	x		3	1.39	.65
13	x	x	x		next year	4—	3.24	1.05
14	x	useful	x	x		3	3.45	.73
15	x	x	meet	x		3	2.87	.89
16	x	am like	-	x		3	2.87	.92
17	x	x	x	x		4	3.45	.72
18	parents	x	x	x	by me	3—	3.60	.65
19	x	x	hear	x		3	2.77	.86
20	Implied	interest	x	x		3	2.49	.88
21	x	helpful	x	x		3	2.54	.99
22	everyone	should	x	f.l.		1	3.03	.95
23	Americans	need	x	f.l.		1	3.06	.97
24	x	helpful	x	x		3	1.86	.85
25	x	easy	x	x		3	2.46	1.09
26	x	-	think	x		3	1.98	.90
27	friends	x	x	x		3	2.43	.92
28	x	x	x	x		4	3.36	.75
29	x	use	-	x	out of class	3—	2.22	.82
30	x	x	read	x	on my own	3—	2.36	.96
31	x	x	x	x	nest year	4—	3.20	1.01
32	Implied	important	x	x		3	3.25	.89

*Abbreviations: x = matches the model, f.l. = foreign language, "—" = not a part of, nor negated by the sentence, "Sum" = the number of x's (followed by a minus sign if additionally qualified).

Semantic bipolar categories were used in comparing each statement to the model:

- (1) self (I) versus other (people),
- (2) like or value versus dislike or not value,
- (3) study versus use;
- (4) Japanese versus foreign language;
- (5) alone versus in class;
- (6) presently versus later.

This technique provided a stable framework with which to compare the items. Incongruent sums indicate items that lack content or semantic validity.

Item means and standard deviations for the total group are also given in Table 7.9. This group of students was fairly positive with regards to most of the items. Although the respondents were not inclined to want to **teach** Japanese (item 12) or to think it would **help in other subjects** (item 24). The correlations between the items and the total score were also determined and items 4, 16, and 25 had low internal consistency (correlations with the total score were less than .20). A final descriptive technique is made by scaling the items. The method of successive intervals was selected as representative (see Fig. 7.3).

Scale Values of Selected Items with Scale Beginning at $-.40$																	
	$-.4$	$-.2$	0	$.2$	$.4$	$.6$	$.8$	1.0	1.2	1.4	1.6	1.8	2.0	2.2	2.4	2.6	2.8
Items			32		30		20			1	31		13	6	28	17	18

FIG. 7.3. Successive interval scaling of foreign language.

A final revised scale is determined by selecting items that (1) provide an approximation to an equal interval scale, (2) satisfy a semantic description and (3) satisfy an item analysis (see Fig. 7.4).

Directions: Select the three statements with which you most nearly agree.

- A. Japanese is one of the most important subjects in the school curriculum. (32)
- B. I'm looking forward to reading Japanese books on my own. (30)
- C. Japanese is one of my most interesting subjects. (20)
- D. I like studying Japanese. (1)
- E. I would like to study more Japanese during the next school year. (31)
- F. I would like to take Japanese again next year. (13)
- G. Japanese is interesting. (6)
- H. I'm glad that I have the opportunity to study Japanese. (28)
- I. I'm glad that Japanese is taught in this school. (17)
- J. My parents are pleased that I'm learning Japanese. (18)

FIG. 7.4. The revised scale of foreign language attitude.

Note: Because of the similarity of wording in items E and F and in items H and I, it may be desirable to either scramble the item order or to eliminate one from each pair.

Discussion

Shaw and Wright (1967) provide a complete text of ordered category scales (see also Robinson, Rusk, & Head, 1969 a, b, c). Check also the bibliography containing references compiled by G. David Garson and the many search engines on the Internet, most preferably google.com, and dialog.com. Kendalls's Tau Correlations can be used to intercorrelate items in categorical rating scales (see SAS PROC CORR K).

Cronbach's Alpha

Cronbach's alpha is most commonly used to determine the reliability of a set of categorical ratings. Alpha reliability is based on the assumptions that item variance is error variance. One definition of reliability is:

$$r_{xx} = S^2_{\text{True}}/S^2_{\text{Total}}.$$

Because True variance is elusive (not known) it is estimated by subtracting estimates of the Error variance from the Total Variance of the test.

$$S^2_{\text{True}} = S^2_{\text{Total}} - S^2_{\text{Error}} \text{ and}$$

$$r_{xx} = S^2_{\text{Total}} - S^2_{\text{Error}}/S^2_{\text{Total}} \text{ or } 1 - S^2_{\text{Error}}/S^2_{\text{Total}}$$

The Sum of the item variances is substituted for the Error variance and a correction for small numbers of k items is applied and the formula is:

$$\alpha = k/k - 1(1 - \sum S^2_{\text{Item}}/S^2_{\text{Total}})$$

When items are scored dichotomously (1,0) then the simple product of (p), the proportion passing and (1-p) or q equals the item's variance. $\sum pq$ is then substituted for the error and KR21 reliability results. Note: $\sum pq = \sum (X - \bar{X})^2/N$.

Programs: SAS Proc Means, alpha, r_{total} and SPSS

One can use SAS to calculate alpha as part of the Proc Means procedure. SPSS has a statistics option which allows the researcher to select variables they wish to constitute a dimensional subset. The objects can be summed and designated as a new variable. This new variable can then be correlated with information such as gender, age, years of education, and other demographics. The correlation of each item with this sum can also be easily determined and a reliability measure found.

This page intentionally left blank

PART III

CLUSTERING

Clustering methods have value because of their simplicity and lack of assumptions. Cluster analysis is a general term for those methods that attempt to separate objects or individuals into groups. The methods covered in this section are ones that the authors have found to be useful and represent the major techniques. These methods include (1) graphic similarity analysis, (2) single and complete linkage clustering, (3) divisive clustering, and (4) K- means iterative clustering. The essential function of these methods is data reduction and description, but they can be useful in hypothesis generation or hypothesis testing by uncovering a meaningful hidden structure.

It is useful to routinely cluster the items in any ordered category instrument. Clustering provides information on the probable dimensionality of the instrument and pictorially identifies items that do not belong.

The typical analysis for clustering is as follows:

1. Objects are identified.
2. Similarities or distances between all objects are obtained.
3. Objects are grouped together based on a measure of distance.
4. A graphical representation of the groups is created.
5. An interpretation is made.

Reverse Scoring for Negative Items

Similarities are often correlations. If correlations are used with negatively worded items, the original scoring of these items should be reversed so that similar items will be correlated positively. The typical correlation between two items such as "I like school" and "School is worthless" would be negatively correlated. Reversing the scoring means that "I like school" and "School is *not* worthless" would be positively correlated. This is an important consideration for clustering methods which convert similarities to distances. With ordered category data and unit weights, a reverse score is equal to $K + 1$ categories minus the subjects response score. If, for a given item, a strongly disagree response is scored as 5, then the reverse value, for $K = 5$ categories, is equal to $(6 - 5)$ or 1.

This page intentionally left blank

8

GRAPHIC SIMILARITY ANALYSIS

The first technique, proposed by Waern (1972), is one of the simplest methods for analyzing similarities among the members of a set of objects. In this method, the magnitudes of the similarity measures are utilized in stepwise fashion. Initially, the experimenter sets absolute or relative standards of association under which he or she will identify and select pairs of objects at or above the standard. Waern cautions conservatism in the selection of standards. Little insight is be gained by making all possible connections.

If the standard chosen for clustering are levels of significance for correlations, the experimenter may choose conventional .01, .05, and .10 significance levels. One may, however, choose cut-off values for each step that include certain small percentages of the pairwise similarity data. For example, steps that include 5% of the highest similarity values at each step might be chosen. That is, steps would be chosen such that the top 5% of all similarity data values would be included in step 1, the second 5% would make up step 2, and so on.

Graphing Ability and Achievement

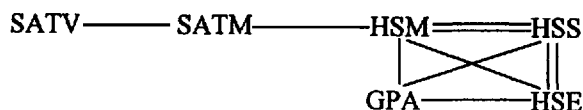
Table 8.1 consists of Pearson correlation coefficients between ability and achievement scores based on N = 224 (Moore, 1994). They consist of high school seniors scores on the Scholastic Achievement Test (Verbal and Math), their grade point average (GPA) and high school math

Table 8.1

Correlations Between Ability and Achievement Scores

	GPA	SATM	SATV	HSM	HSS	HSE
GPA						
SATM	.25					
SATV	.11	.46				
HSM	.44	.45	.22			
HSS	.33	.24	.26	.58		
HSE	.29	.11	.24	.45	.58	

(HSM), English (HSE) and science (HSS) grades. Because N is so large almost all the correlation coefficients are statistically significant at $p < .01$. A decision is made, therefore, to connect tests with two lines which are correlated $r > .50$ and with one line for test correlated $r > .40$ but $r < .50$. The graph is as follows:



The graph illustrates, with the exception of math grades, that high school grades and GPA are marginally related to scores on the SAT.

Generally conventional levels of statistical significance .01, .05, .10 are used. A search for all pairwise correlations that reach the .01 random chance level ($p < .01$) is made as a first step. These selections are drawn in the two-dimensional space of a plain sheet of paper and connected by two heavy black lines. Next the correlations at the .05 level are placed, using single lines, and finally those at the .10 level, using dotted lines. The length of the lines is not generally used as an indication of similarity. Should negative correlations occur an **O** is placed on the connecting line. For example, if A correlates with B $-.78$, $p < .05$. A — **O** — B.

Some subjectivity is necessary in this analysis but the results can be revealing and rewarding because subtle connections between clusters and chains may be lost in more conventional methods such as factor analysis and multidimensional scaling. The technique is particularly useful in interpreting the results from other more technical methods.

Graphing Letter Similarity

In this illustration, connecting lines indicate known similarities. Above and below the “ground” letters (a, c, o p, b, h, n, etc.) are “figural” letters. The left oriented letters (q, d, j, i, and l) are together at the top of Figure 8.1 and the linearly dominated (z, x, y, and v) are at the bottom. (Dunn-Rankin, 1990) The letters have also been arranged according to familiar feature patterns as found by Dunn-Rankin (1968), Kuennapas and Janson (1969) and Bouma (1971).

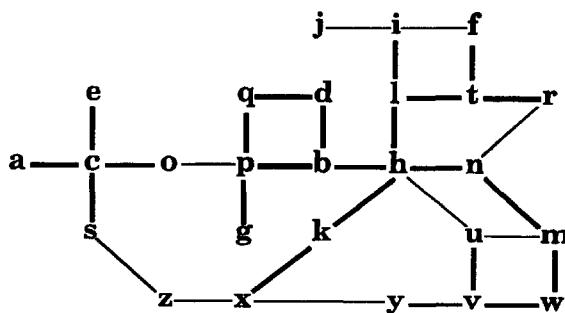


FIG. 8.1. Graphic similarity of lowercase English letters. Dark lines indicate scale values < 20 and light lines indicate values < 30 , where the range is 0 to 100.

The lowercase letters are shown in standard form. The letters are arranged to accomodate the similarities found in several studies. The letters with large almost closed areas are located in the middle of the array. Linear or figural letters are located on the edges. One could conceivably roll the array end to end and edge to edge to form a multidimensional donut or torus of letters.

Graphic Analysis of Word Similarity

Dunn-Rankin (1995) analyzed 50 graduate students' reponses to the free clustering of 22 words following the instructions: "Group similar words together." In Fig. 8.2, a network graph of the percent overlap similarity measures is presented for all the subjects. It is easy to delineate the relative importance of graphic, phonetic, or semantic characteristics for specific words. For example, fool and food are seen as graphically and phonetically similar, whereas fool and fellow have semantic similarity. Students were later easily clustered into groups which preferred either the semantic, phonetic, or graphic dimensions of similarity.

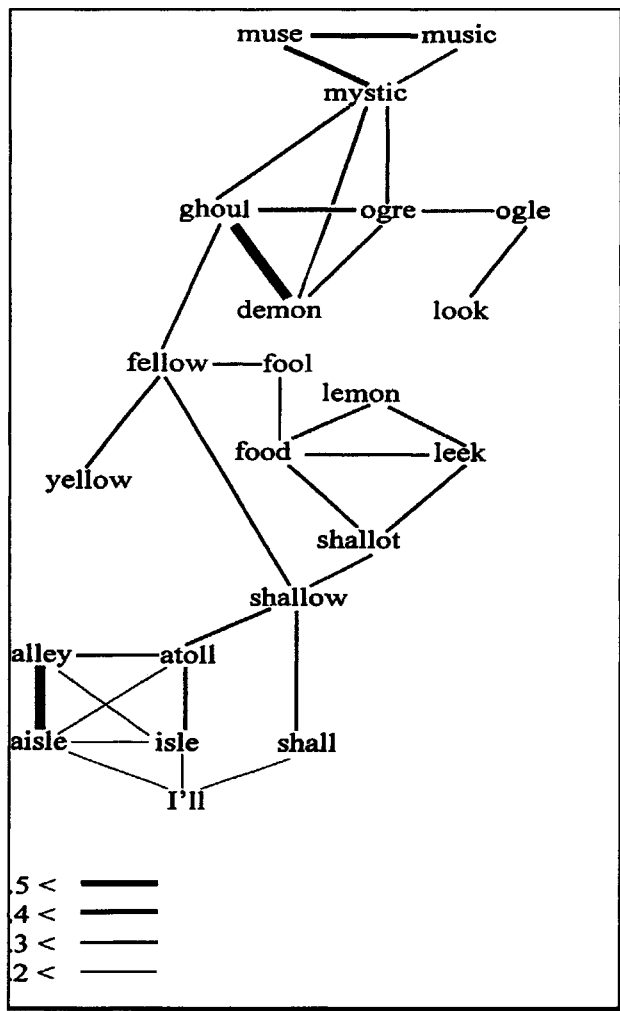
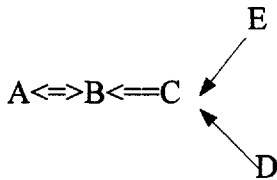


FIG. 8.2. A network graph of the percent overlap measures for 50 graduate students on

Elementary Linkage Analysis

Elementary linkage analysis (McQuitty, 1957) is a simple method of clustering. It can be used to cluster any objects (people or items) that have distinctive characteristics of similarity. Advantages of elementary linkage analysis are speed, objectivity and its provision for investigating a particular theoretical position. A 15-variable matrix can be analyzed into objectively determined "types" in 5 to 10 minutes. Furthermore, all elementary linkage analysis operations require only pencil and paper. (McQuitty's method has been programmed and is available in SAS PROC Cluster.) The steps are:

1. Underline the highest (absolute) entry in each **column** of the matrix.
2. Select the highest entry in the matrix. Write the variable code on a piece of paper with reciprocal arrows; for example: $A \rightleftharpoons B$. Call this the first Type or Type I.
3. Select all those variables (objects or subjects) that are most like members of the first Type by reading only the **rows** containing the first two variables (A and B) and selecting previously **underlined** entries in these rows. Write down these variable codes and connect them to the related variable by a single arrow; for example: $A \rightleftharpoons B \leftrightsquigarrow C$. Call these variables (C) first Cousins.
4. Select all those variables that are like the first-Cousin variables. Write them down with connecting arrows; for example:



Call these variables (D and E) second cousins.

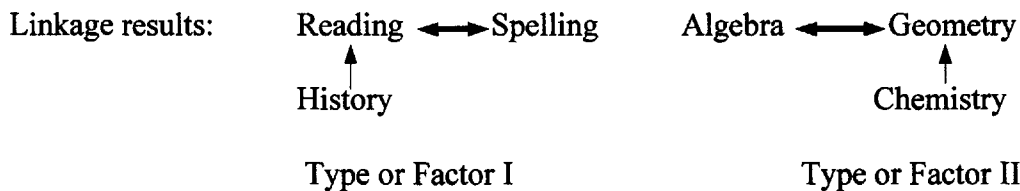
5. Search for higher order cousins until no more exist.
6. Exclude all variables already classified and repeat steps two to five. Increment the type number by one for each time through the steps.
7. Repeat these steps until all variables are classified.

Linkage Analysis of Test Scores

A linkage analysis is illustrated using the data of Table 8.2. The data set consists of correlations between achievement test scores of high school students. First the highest values were underlined in each **column**. Then the correlation (.72) between spelling (S) and reading (R) is initially selected as Type I ($R \rightleftharpoons S$). A review of the Reading and Spelling **rows** reveals one other underlined correlation (.63) between History (H) and Reading (R). History becomes a first cousin. As no other underlined correlation exists in these two rows, a search is made for the next highest correlation not already utilized. This occurs between Algebra and Geometry where ($r = .64$). This initiates the second type, to which Chemistry is attached. The similarity of linkage analysis to factor analysis has allowed researchers to sometimes equate Types with Factors. (see Factor Analysis p. 149).

Table 8.2
Intercorrelations of Test Scores of High School Students

	Reading	Spelling	Algebra	Geometry	Chemistry	History
Reading		<u>.72</u>	.12	.27	.20	<u>.63</u>
Spelling	<u>.72</u>		.16	.40	.12	.56
Algebra	.12	.16		<u>.64</u>	.48	.10
Geometry	.27	.40	<u>.64</u>		<u>.50</u>	.21
Chemistry	.20	.12	.48	.50		.14
History	.63	.56	.10	.21	.14	



Discussion

Graphic methods can be quickly applied to any form of proximity matrix. Proximities can be similarities like correlations or distances. In the three examples presented, Graphic Similarity analysis was used on:

1. Ability and achievement. The size of r was used as a standard of similarity.
2. Letter similarity. The size of the scale values was used as a standard.
3. Word similarity. The percentage of overlap was used as a standard.

Nearly every computer center has packaged programs that routinely calculate correlations. The SAS system PROC CORR can be called on to print only the highest correlations with each variable. For example:

```
PROC CORR best = 10;
Var.....;
```

The results will then display each variable with the ten variables most highly correlated and printed in order. This makes selection much easier than perusing a total matrix of r 's. It is possible to write a program to print only values of a certain size, relative size, or significance. This would aid the graphic methods in that the selection of relevant variables would be easier. Although McQuitty's method is done easily by hand, a computer solution is available in SAS PROC Cluster Mcquitty.

This page intentionally left blank

9

SUCCESSIVE COMBINING

Agglomerative methods of clustering begin with the calculation of a distance matrix. Similarities such as correlations are converted to distances by subtracting each value from a constant such as 1.00. Objects which are most similar and therefore not far apart are grouped together. The number of groups (each object is a group of one to start with) is reduced in stepwise fashion by forming new groups of similar objects at each step. The process is continued until all the objects are formed into a single group. The graphic representation of this process is called a tree or dendogram.

Ward's Mimimum Variance Method

Ward (1963) uses the "**sum of squares**" as the basis for inclusion of objects in a group. The sum of squares, a measure of variation, is given by

$$SS = \sum (X - \bar{X})^2 = \sum X^2 - (\sum X)^2 / N$$

where X is the variable and N is the number of objects. At each stage of the clustering, SS is calculated for every possible pair of objects and the pair is chosen which provides the minimum sum of squares. Suppose paired **distances** between six lowercase letters are as listed in Table 9.1:

Table 9.1
Distances Between Letters

		1	2	3	4	5	6
		x	k	h	q	p	g
1	x	0					
2	k	2	0				
3	h	4	2	0			
4	q	8	6	4	0		
5	p	10	8	6	2	0	
6	g	14	12	10	6	4	0

In this matrix, direct estimates of the distances between lower case letters of the English alphabet are presented. It can be seen that the distance between any letter and itself is zero. Therefore the calculation of the SS between letters x and k is as follows:

	d	d ²
(1) x	0	0
(2) k	2	4
Σ	2	4

$$SS = \Sigma d^2 - (\Sigma d)^2/N$$

$$= 4 - (2)(2)/2$$

$$= 2$$

Initially each object (letter) is considered a group consisting of a single member. Next all possible pairwise groups are formed and the pair that has the smallest SS forms the first group. The process is repeated with $K - 1$ groups until all the objects are in a single group. Table 9.2 provides the sum of squares for all possible (15) pairs of letters.

Table 9.2
SS for Six Letters

	1	2	3	4	5	6
1						
2	2					
3	8	2				
4	32	18	8			
5	50	32	18	2		
6	98	72	50	18	8	

Because the pairs 1-2, 2-3, and 4-5 have equally small sums of squares, the first clustering is arbitrary. Objects 1 and 2 are chosen simply because they are first.

By calling 1 and 2 a new group A (1&2) the six original objects or groups have been reduced to five:

A (1&2), 3, 4, 5, 6,

In the second stage of clustering there are (5 choose 2) or 10 possible pairings:

Pairs	A-3	A-4	A-5	A-6	3-4	3-5	3-6	4-5	4-6	5-6
SS	8	35	56	115	8	18	25	2	18	8

Using the table of distances the SS for A-3 is calculated $(0 - 2)^2 + (2 - 2)^2 + (4 - 2)^2 = 8$, where $(0 + 2 + 4)/3 = 2$ is the mean.

Now 4 and 5 are grouped (SS = 2). Calling this new group B (4 & 5), four groups remain:

A (1 & 2) 3 B (4 & 5) 6
There are six pairs in the third stage.

Pairs	A-3	A-B	A-6	B-3	3-6	B-6
SS	8	68	115	19	25	9

Now A and 3 are combined. Calling this new group C (1 & 2 & 3) we have groups

C	B	6
---	---	---

Next the sums of squares for these three groups are calculated.

Pairs	C-B	C-6	B-6
SS	35	115	9

Finally, B and 6 are combined. Calling the new group D, only groups C and D are left to combine and this combination is called E. This analysis can be summarized in a dendrogram as follows:

Because the increase in SS to combine C and D is relatively large the experimenter may wish to stop clustering after this step. Any hidden structure revealed by the clustering should be visually

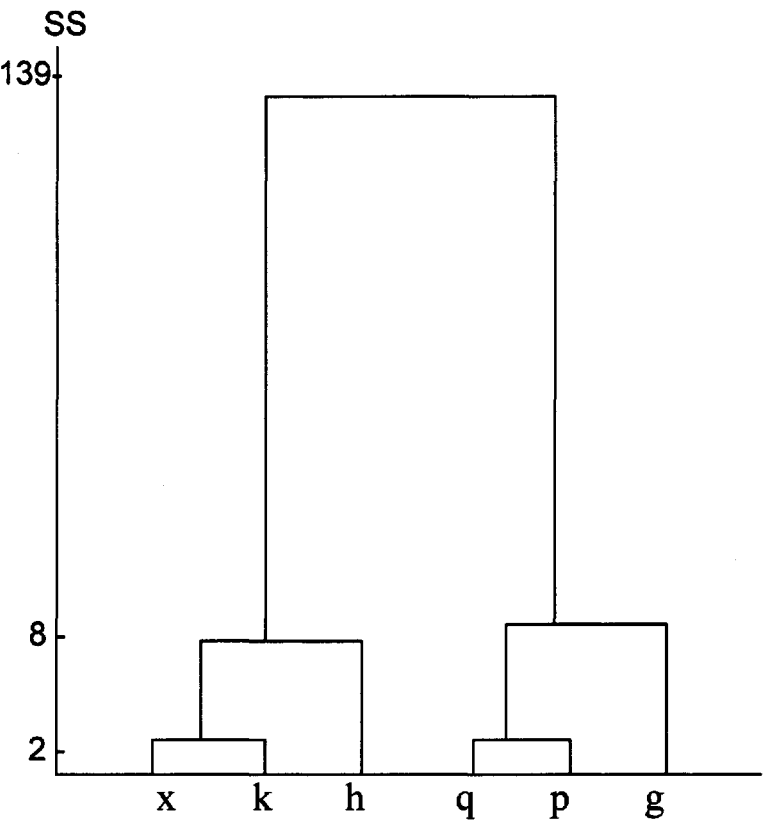


FIG. 9.1. A dendrogram showing the hierarchical clustering of lower case letters. Note that two clusters differ markedly. Cluster x, k, h has an angular component whereas Cluster q, p, g has a circular component with a descender.

stop clustering after this step. Any hidden structure revealed by the clustering should be visually apparent. In this example it is easy to see that (x, k, h) differ from (q, p, g) . A somewhat more objective means of finding the right number of clusters is to divide the total sums of squares by the number of objects and use this average SS value as an interval on the y axis. The clusters are equally spaced on the x axis. A plot of clusters versus. SS will reveal when it no longer is useful to continue clustering (see p. 131).

When starting with a similarity or distance matrix it can be shown (Anderberg, 1973) that the pairwise distance measures are proportional to the calculated sums of squares between all pairs of objects. The entries in the distance matrix (without summing the squares), therefore, may be taken as the *first step* in Ward's method. Ward (1980) relates that for job classification and task analysis, percentage of overlap is used more often than correlation as a measure of similarity. In order to do Ward's minimum variance method use:

SAS PROC Cluster Method = Ward.

Grouping Students on Reward Preference

Dunn-Rankin, Shimizu, and King (1969) studied the reward preference of fifth and sixth-grade children. They had students respond to a paired comparison task involving five kinds of reward using a Reward Preference Inventory (RPI): Adult Approval (A), Competitive Rewards (C), Peer Approval (P), Independence Rewards (I), and Consumable (eatable) Rewards (E). A child would be asked, for example, whether he or she would rather have an **Excellent** on his or her paper (A) or some **candy** (E).

As shown in Fig. 9.2 the 10 possible pairs of rewards were repeated four times using parallel items. The summed votes provided a reward preference profile for each subject. The profiles were intercorrelated and the correlation values were used in Ward's (1963) method to cluster the subjects. A final selection of four groups was made and these groups are presented in Table 9.3.

RPI

Directions: Which one of each pair do you like best? Circle the Letter to the right of each statement you like best

- | | | |
|----|--|--------|
| 1. | Teacher writes "100" on your paper.
Be first to finish your work. | A
C |
| 2. | Package of bubble gum.
Students ask you to be on their team. | E
P |
| 3. | Be free to do what you like.
Teacher writes "100" on your paper. | I
A |
| 4. | Students ask you to be on their team.
Be first to finish your work. | P
C |
| 5. | Be free to do what you like.
A package of bubble gum. | I
E |
| 6. | Teacher writes "100" on your paper.
Students ask you to be on their team. | A
P |

- | | | |
|-----|---|--------|
| 7. | Be first to finish your work.
Be free to do what you like. | C
I |
| 8. | A package of bubble gum.
Teacher writes "100" on your paper. | E
A |
| 9. | Students ask you to be on their team.
Be free to do what you like. | P
I |
| 10. | Be first to finish your work.
A package of bubble gum. | C
E |
| 11. | Be the only one that can answer a question.
Candy | C
E |
| 12. | Friends ask you to sit with them.
Be free to go outside. | P
I |
| 13. | Candy.
Teacher writes "A" on your paper. | E
A |
| 14. | Be the only one that can answer a question.
Be free to go outside. | C
I |
| 15. | Teacher writes "A" on your paper.
Friends ask you to sit with them. | A
P |
| 16. | Be free to go outside.
Candy. | I
E |
| 17. | Friends ask you to sit with them.
Be the only one that can answer a question. | P
C |
| 18. | Be free to go outside.
Teacher writes "A" on your paper. | I
A |
| 19. | Candy.
Friends ask you to sit with them. | E
P |
| 20. | Teacher writes "A" on your paper.
Be the only one that can answer a question. | A
C |
| 21. | Have only your paper shown to the class.
Teacher writes "very good" on your paper. | C
A |
| 22. | Classmates ask you to be a class leader.
Ice Cream. | P
E |
| 23. | Teacher writes "very good" on your paper.
Be free to play outside. | A
I |
| 24. | Have only your paper shown to the class.
Classmates ask you to be class leader. | C
P |
| 25. | Ice cream.
Be free to play outside. | E
I |
| 26. | Classmates ask you to be class leader.
Teacher writes "very good" on your paper. | P
A |

27.	Be free to play outside. Have only your paper shown to the class.	I C
28.	Teacher writes "very good" on your paper. Ice cream.	A E
29.	Be free to play outside. Classmates ask you to be class leader.	I P
30.	Ice cream. Have only your paper shown to the class.	E C
31.	A soft drink. Have your paper be the best in the class	E C
32.	Be free to work on something you like. Friends ask you to work with them.	I P
33.	Teacher writes "excellent" on your paper. A soft drink.	A E
34.	Be free to work on something you like. Have your paper be the best in the class.	I C
35.	Friends ask you to work with them. Teacher writes "excellent" on your paper.	P A
36.	A soft drink. Be free to work on something you like.	E I
37.	Have your paper be the best in the class. Friends ask you to work with them.	C P
38.	Teacher writes "excellent" on your paper. Be free to work on something you like.	A I
39.	Friends ask you to work with them. A soft drink.	P E
40.	Have your paper be the best in the class. Teacher writes "excellent" on your paper.	C A

FIG. 9.2. Reward Preference Inventory. Counts of (A) Adult approval, (C) Competitive Rewards, (E) Eatables, (P) Peer Approval and (I) Independence create a profile.

Table 9.3
Major Preference Profile Groups of 5th and 6th Grade Children
Related Behavioral and Ethnic Variables

Profile					Variables						
A	C	P	I	E	Sex	Ethnic Origin	Grade	SES	IQ	Step Math	Step Reading
Group 1											
16	5	8	17	14	0	0	6	3	133	258	259
16	9	8	14	13	1	N	5	6	131	233	252
11	8	11	18	12	1	0	5	6	141	257	264
9	4	11	20	16	1	N	6	6	127	267	279
10	4	11	20	15	1	N	6	2	139	261	263
14	6	13	14	13	1	N	6	6	115	244	258
12	8	10	12	18	1	0	5	3	137	254	259
11	6	7	16	20	1	0	6	3	129	265	267
12	5	7	16	20	1	0	6	5	140	252	283
16	12	8	20	4	0	0	5	6	137	253	267
14	11	5	20	10	1	N	6	0	121	251	253

Group 2											
16	4'	20	12	8	0	0	5	5	134	263	273
15	6	20	13	6	0	0	5	6	141	251	261
11	9	20	16	4	0	0	5	5	152	258	276
12	8	20	16	4	0	0	6	2	135	266	296
12	8	19	17	4	0	0	6	5	133	249	251
14	8	19	15	4	1	0	6	6	138	263	287
12	7	18	18	5	0	0	6	6	141	269	291
18	6	14	13	9	0	0	5	2	131	244	260
16	6	15	17	6	0	0	5	4	110	249	267
15	10	12	15	8	1	0	5	3	120	240	242
17	9	11	17	6	0	0	6	6	125	251	259
18	7	11	16	8	1	0	6	5	131	259	266
Group 3											
17	13	17	9	4	0	0	6	6	109	244	248
17	11	17	9	6	0	0	5	3	129	256	261
15	13	20	8	4	1	0	5	5	131	261	273
20	10	15	11	4	0	N	5	3	130	257	255
19	11	14	11	5	1	N	5	6	125	245	267
19	11	15	11	4	0	N	6	4	107	245	256
18	11	16	11	4	0	0	5	5	126	253	273
18	12	16	10	4	0	0	6	5	88	245	251
17	12	15	12	4	1	0	5	3	111	254	262
17	10	15	13	5	0	N	6	2	123	263	271
20	10	13	12	5	0	0	6	3	99	251	257
19	11	13	13	4	0	0	6	0	100	258	252
19	10	12	11	8	0	0	5	6	125	230	255
10	8	15	10	8	0	N	6	5	99	266	255
Group 4											
20	16	12	7	5	0	0	5	3	121	239	242
19	16	13	7	5	1	0	6	5	100	243	254
19	17	12	6	6	0	0	5	1	118	243	262
20	16	12	8	4	1	0	5	6	120	239	256
20	16	12	8	4	1	0	6	3	102	250	260
19	16	12	9	4	0	0	6	3	126	249	261
19	16	13	8	4	0	N	5	5	108	240	255
19	17	12	8	4	1	0	6	3	112	258	267
17	18	12	8	5	0	N	5	6	127	245	242
20	13	15	8	4	0	0	5	3	115	245	261
20	13	14	9	4	0	0	6	6	146	261	257
19	15	14	8	4	0	0	6	5	112	243	252
20	14	14	8	4	1	N	6	4	126	240	247
18	14	14	10	4	0	N	6	5	139	270	271
18	14	14	7	7	1	N	5	5	113	244	259
20	14	11	11	4	1	N	6	0	111	263	266
18	14	12	11	5	0	0	5	5	119	257	260
19	15	11	11	4	1	0	6	6	120	249	267
19	14	9	14	4	1	0	5	6	117	247	262
16	15	10	13	6	1	0	6	3	112	249	255
19	17	9	11	4	0	N	5	6	127	252	267
17	13	12	14	4	0	N	5	6	134	251	263
16	14	13	13	4	0	0	6	5	123	266	264

Table 9.4
Summary Table

Group	Average Profile		P	I	E	N	Male	In Grade		6	SES	Step	Step	Math	Read
	A	C						Orient							
1	13	7	9	17	14	11	9	6	7	4.2		131.8	254.1	264.0	
2	15	7	17	15	6	12	3	12	6	4.6		132.6	255.2	269.1	
3	18	11	15	11	5	14	3	9	7	4.0		114.4	252.0	259.7	
4	19	15	12	9	5	24	10	15	12	4.4		119.0	249.2	257.8	
Total						61	25	42	32	4.3		123.0	251.9	261.6	

Table 9.4 illustrates that reward preference is related to a variety of variables in different ways. The mean profiles for each group illustrate that ability and achievement, judged by SCAT (School and College Ability) and STEP (Sequential Tests of Educational Progress) tests, are related to high scores for independence rewards. Lower achievement is related to a professed desire for adult approval. A theory of relative satiation for these rewards is postulated.

CD-ROM and SAS Clustering Example

PROC Cluster in SAS starts with a distance matrix. If the data are similarities they must be converted to distances prior to the analysis. In this example, the initial data come from the free clustering of 11 words all beginning with the letter **a**. First we use program **perover.exe** on the CD-ROM to convert the group numbers to distances. Note this is sample data for illustration purposes only.

Words.cfg: A configuration file is formed

Similar Words Beginning With a	Title
14 11	Number of Subjects and Number of variables
Words.dat	Input data file
Words.out	Output data file

Words.dat: A data file is also needed

1 2 3 2 4 3 1 1 5 6 6	
1 2 3 2 2 1 1 1 3 2 2	
1 2 1 2 2 3 1 1 3 3 3	
1 2 3 4 4 1 5 6 7 8 8	
1 2 3 4 4 1 5 6 7 8 8	
1 2 3 3 4 5 1 6 7 8 8	
1 2 2 2 2 3 1 1 2 2 2	
1 2 3 4 5 6 7 7 8 9 9	
1 2 3 2 4 5 1 6 4 4 4	
1 2 3 4 5 2 1 1 2 6 6	
1 2 3 4 2 3 1 1 3 3 3	
1 2 3 2 4 5 1 1 6 7 5	
1 2 3 2 4 5 1 1 6 7 7	
1 2 3 2 2 3 1 1 3 2 3	
	}Data: group membership numbers
	End of file carriage return

Words.out: The Output File

Words Beginning With a
subjects = 14 variables = 11
raw data

1.	2.	3.	2.	4.	3.	1.	1.	5.	6.	6.
1.	2.	3.	2.	2.	1.	1.	1.	3.	2.	2.
1.	2.	1.	2.	2.	3.	1.	1.	3.	3.	3.
1.	2.	3.	4.	4.	1.	5.	6.	7.	8.	8.
1.	2.	3.	4.	4.	1.	5.	6.	7.	8.	8.
1.	2.	3.	3.	4.	5.	1.	6.	7.	8.	8.
1.	2.	2.	2.	2.	3.	1.	1.	2.	2.	2.
1.	2.	3.	4.	5.	6.	7.	7.	8.	9.	9.
1.	2.	3.	2.	4.	5.	1.	6.	4.	4.	4.
1.	2.	3.	4.	5.	2.	1.	1.	2.	6.	6.
1.	2.	3.	4.	2.	3.	1.	1.	3.	3.	3.
1.	2.	3.	2.	4.	5.	1.	1.	6.	7.	5.
1.	2.	3.	2.	4.	5.	1.	1.	6.	7.	7.
1.	2.	3.	2.	2.	3.	1.	1.	3.	2.	3.

A distance matrix between the 11 words beginning with a is also part of the output by **PEROVER** which measures the percent overlap from free clustering.

Distance (difference) matrix

```

0.000 1.000 0.929 1.000 1.000 0.786 0.214 0.357 1.000 1.000 1.000
1.000 0.000 0.929 0.429 0.643 0.929 1.000 1.000 0.857 0.786 0.857
0.929 0.929 0.000 0.857 0.929 0.786 0.929 0.929 0.714 0.857 0.786
1.000 0.429 0.857 0.000 0.571 1.000 1.000 1.000 0.929 0.786 0.857
1.000 0.643 0.929 0.571 0.000 1.000 1.000 1.000 0.857 0.714 0.786
0.786 0.929 0.786 1.000 1.000 0.000 0.929 0.929 0.714 0.857 0.714
0.214 1.000 0.929 1.000 1.000 0.929 0.000 0.286 1.000 1.000 1.000
0.357 1.000 0.929 1.000 1.000 0.929 0.286 0.000 1.000 1.000 1.000
1.000 0.857 0.714 0.929 0.857 0.714 1.000 1.000 0.000 0.714 0.643
1.000 0.786 0.857 0.786 0.714 0.857 1.000 1.000 0.714 0.000 0.143
1.000 0.857 0.786 0.857 0.786 0.714 1.000 1.000 0.643 0.143 0.000

```

Ward's hierarchical grouping methodology can be found as one of the methods in the Statistical Analysis System (SAS, 2000) under the procedure **PROC Cluster**. The template example provided below uses the results of **Words.out** as the input.

SAS file

data words;	Defines data file
input a admits aged almost aiming and as at areas army away;	List variables
lines;	Starts data
0.000 1.000 0.929 1.000 1.000 0.786 0.214 0.357 1.000 1.000 1.000	
1.000 0.000 0.929 0.429 0.643 0.929 1.000 1.000 0.857 0.786 0.857	
0.929 0.929 0.000 0.857 0.929 0.786 0.929 0.929 0.714 0.857 0.786	
1.000 0.429 0.857 0.000 0.571 1.000 1.000 1.000 0.929 0.786 0.857	
1.000 0.643 0.929 0.571 0.000 1.000 1.000 1.000 0.857 0.714 0.786	} Data lines
0.786 0.929 0.786 1.000 1.000 0.000 0.929 0.929 0.714 0.857 0.714	
0.214 1.000 0.929 1.000 1.000 0.929 0.000 0.286 1.000 1.000 1.000	
0.357 1.000 0.929 1.000 1.000 0.929 0.286 0.000 1.000 1.000 1.000	
1.000 0.857 0.714 0.929 0.857 0.714 1.000 1.000 0.000 0.714 0.643	
1.000 0.786 0.857 0.786 0.714 0.857 1.000 1.000 0.714 0.000 0.143	
1.000 0.857 0.786 0.857 0.786 0.714 1.000 1.000 0.643 0.143 0.000	
;	
proc cluster method=ward	Starts Ward Clustering
outtree=tree;	SAS output data tree
proc tree;	Proc tree on "tree"
run;	

Results of this CLUSTER procedure provides a Cluster History and a measure of the sums of squares determined as each cluster is joined.

Cluster History

NCL*	Clusters Joined		Freq	SPRSQ	RSQ
11	(None Joined)				
10	OB10	OB11	2	0.0017	.998
9	OB7	OB8	2	0.0064	.992
8	OB1	CL9	3	0.0094	.983
7	OB2	OB4	2	0.0142	.968
6	OB3	OB9	2	0.0411	.927
5	CL7	OB5	3	0.0453	.882
4	CL6	OB6	3	0.0598	.822
3	CL4	CL10	5	0.1570	.665
2	CL5	CL3	8	0.1920	.473
1	CL8	CL2	11	0.4731	.000

*NCL is the number of clusters (11 objects to start with), OB is object and Cl is cluster. Freq is the number of objects contributing to each cluster. SPRSQ (SemiPartial RSQ) is a measure of the variance used to join the objects in the particular cluster. 1 – RSQ is the amount of variance unaccounted for.

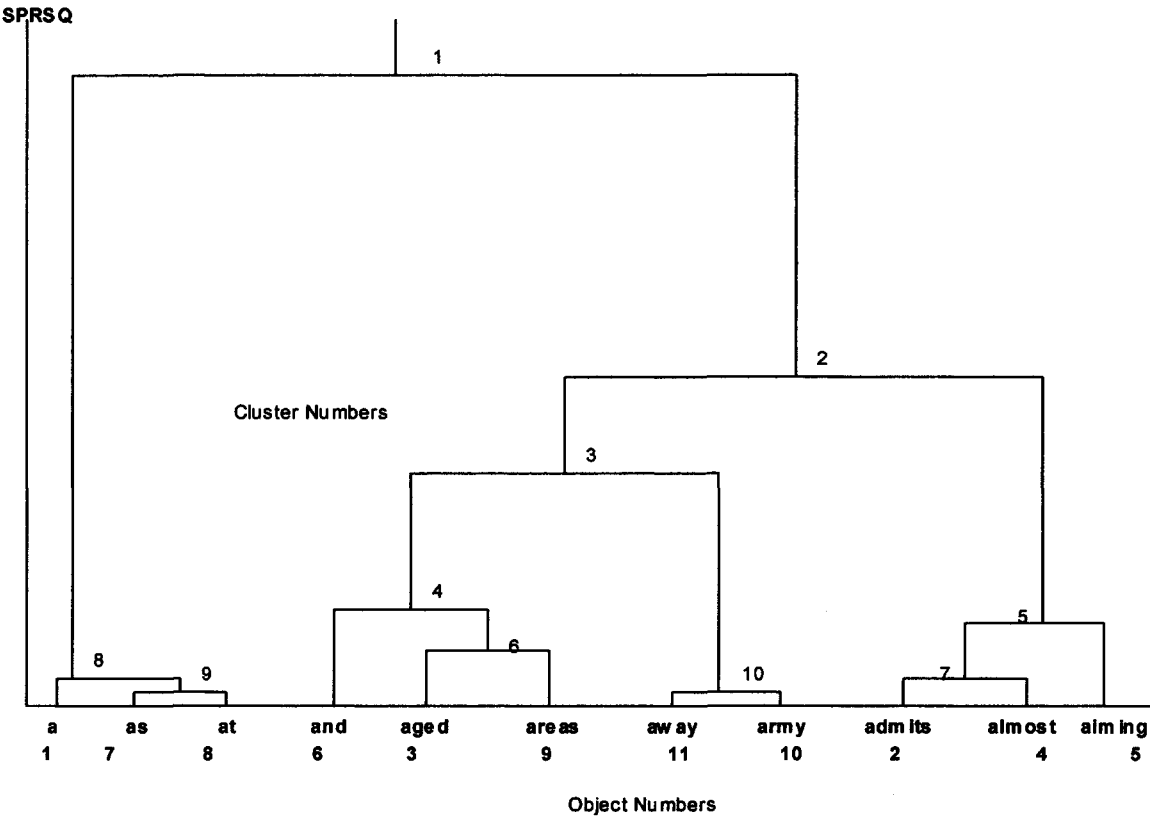


FIG. 9.3. The words and their object numbers are given at the bottom of the graph and the number of the clusters are given at the nodes.

Discussion

The clustering reveals that when words begin with the same letter and have few semantic similarities then judges use the graphic feature of **length** as an important dimension in deciding which words to group together as similar. This is evidenced by the large amount of variance necessary to group the very small words (**a**, **as**, **at**) with the other words. The words **admits**, **almost** and **aiming** are the longest words and each has two or more ascenders. The middle cluster consists of middle sized words.

In chapter 14 (p. 183), (MDS), another look is given to the data in which the clustering is superimposed on a two dimensional plot of the words.

Because Ward's method provides a "variance accounted for" measure at each hierarchical step, a diagram can be drawn to reflect the ease or difficulty in joining any two clusters. The final clusters are ordered sequentially on the horizontal axis from left to right. The total Sums of the Squares (SS) is equally divided by the number of clusters and projected on the vertical axis. A smooth curve is plotted at the height of the SS for each cluster. A rule of thumb is that when the rise and run are equal further clustering is unnecessary (see p. 124).

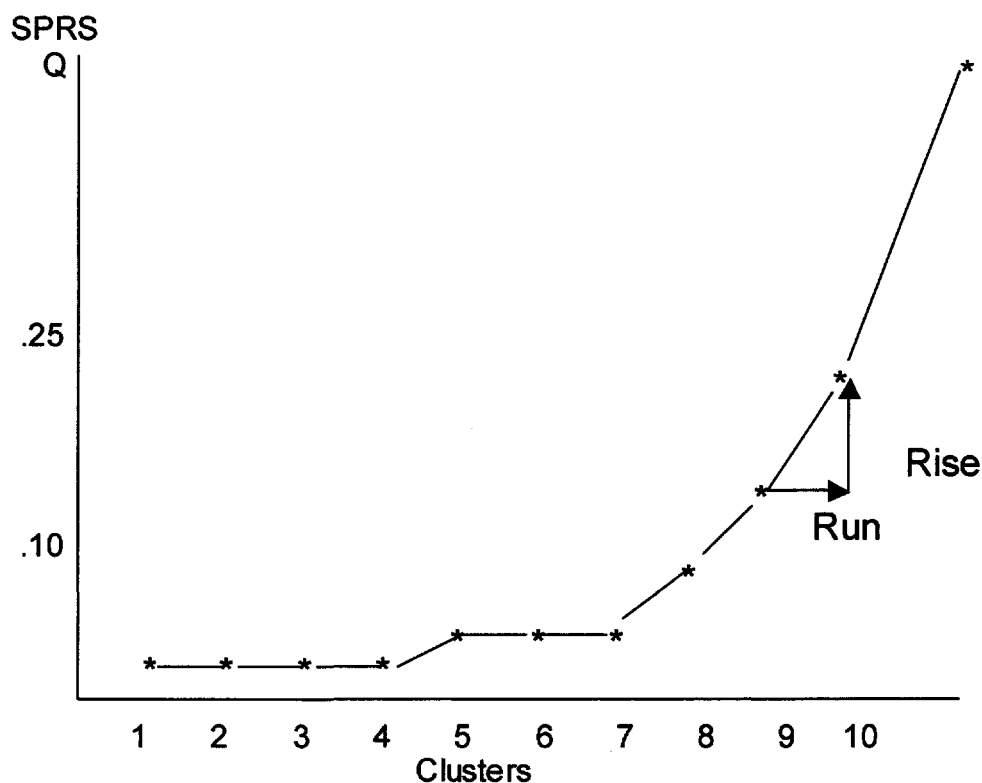


FIG. 9.4. Plot of clusters against variance accounted for. When Rise = Run stop clustering.

Johnson's Nonmetric Single and Complete Link Clustering

In Johnson's (1967) nonmetric clustering method similarities are converted to distances. Distances are either measured to the closest member in a cluster or the farthest member in a cluster. As in most clustering methods, the basic steps are as follows:

1. Gather and establish a data matrix.
2. Calculate a distance matrix between pairs of objects.
3. Join the closest objects into a cluster.
4. Join the next closest pair of objects (using the closest or farthest member of a cluster to represent that cluster).
5. Continue until all objects are in a single cluster.

Suppose that five objects are to be grouped hierarchically and the matrix of distances (d_{ij}) between the objects has been calculated as shown in Table 9.4.

Table 9.4
Distance Matrix

	1	2	3	4	5
1					
2	4				
3	6	3			
4	8	7	2		
5	10	5	6	<u>1</u>	

Objects 4 and 5 should be clustered first, as the distance between these two objects is the smallest in the table ($d_{4:5} = 1$). Thus objects 4 and 5 are united into a single group.

Under the complete linkage method sometimes known as the "farthest neighbor" method the **maximum** distances between the **elements** of the new group and each of the remaining objects are obtained.

$$\begin{aligned}
 d_{(4:5)1} &= \max (d_{14}, d_{15}) = 10 \quad \text{Comparing } d_{14} = 8 \text{ and } d_{15} = 10, 10 \text{ is maximum} \\
 d_{(4:5)2} &= \max (d_{24}, d_{25}) = 7 \quad \text{Comparing } d_{24} = 7 \text{ and } d_{25} = 5, 7 \text{ is maximum} \\
 d_{(4:5)3} &= \max (d_{34}, d_{35}) = 6 \quad \text{Comparing } d_{34} = 2 \text{ and } d_{35} = 6, 6 \text{ is maximum}
 \end{aligned}$$

These maximum distance values are now entered into a reduced matrix of distances as shown in Table 9.5. The new matrix is reduced by eliminating the minimum values of 4 and 5 and keeping the maximum values. (Note, it is also possible to retain the minimum distances and a different or "nearest neighbor" analysis can be made).

Table 9.5
Reduced Distance Matrix

	1	2	3	(4:5)
1				
2	4			
3	6	<u>3</u>		
(4:5)	10	7	6	

Table 9.5 is now examined for its smallest entry, which occurs between objects 2 and 3 or a value of 3. These objects are united into a single group and the maximum values from this new entry are determined.

$$d_{(2:3)1} = \max(d_{12}, d_{13}) = d_{13} \text{ or } 6$$

$$d_{(2:3)(4:5)} = \max(d_{2(4:5)}, d_{3(4:5)}) = d_{2(4:5)} \text{ or } 7$$

These clusters are now placed in the third reduced matrix, shown in Table 9.6.

Table 9.6
Reduced Distance Matrix

	1	(2:3)	(4:5)
1			
(2:3)	<u>6</u>		
(4:5)	10	7	

The smallest entry in this matrix is $d_{(2:3)1}$ or 6. When these two groups are united only two entries are left to finally unite. A dendrogram of the process is given in Fig. 9.5.

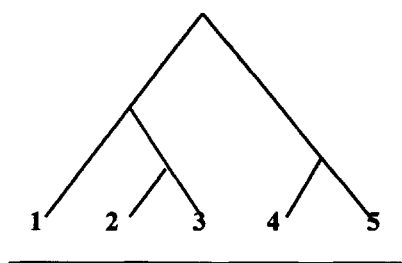


FIG. 9.5. Dendrogram of complete linkage clustering

Clustering the WISC Tests with HICLUS

In the single-link method, the smallest or **minimum** distances to the potential entries of the reduced matrix is used instead of the maximum, otherwise the procedure is the same. The single-link or connectedness method has a strong theoretical rationale in the biological sciences and is useful in determining items that belong to a subtest or factor. Practice shows that the complete-link or diameter method is more effective with social science data. Complete link creates more discrete clusters. Johnson's (1967) nonmetric method, **HICLUS**, is detailed in the CD-ROM. An example is given using the complete link method or diameter method to cluster the correlations between the Wechsler Intelligence Scales for Children (WISC) Table 9.7. The WISC is proposed to have two intelligence measures, a verbal score and a performance measure. The clustering dendrogram below supports two measures of intelligence. The verbal and spatial subtests are clearly delineated.

Table 9.7
Intercorrelations Among Wechsler Subtests

	1	2	3	4	5	6	7	8	9	10	11
information											
vocabulary	0.60										
arithmetic	0.58	0.49									
similarities	0.53	0.44	0.46								
comprehension	0.60	0.57	0.51	0.55							
sentences	0.52	0.46	0.51	0.51	0.53						
animal house	0.41	0.36	0.42	0.31	0.34	0.36					
picture	0.47	0.45	0.42	0.36	0.42	0.35	0.38				
mazes	0.37	0.35	0.41	0.28	0.33	0.30	0.36	0.44			
geometric design	0.40	0.35	0.47	0.30	0.36	0.34	0.43	0.42	0.48		
block design	0.43	0.38	0.50	0.35	0.39	0.38	0.38	0.45	0.46	0.48	

The lower triangular matrix, shown as Table 9.7, is copied into a text file and is named **Wisc.dat**. A configuration file is created.

Wisc.cfg

Clustering Wisc Subtests	Title
11 -1	11 variables which are similarities
Wisc.dat	Input file
Wisc.out	Output

The **HICLUS** routine on the CD-ROM, **hiclus.exe**, is called and the data analyzed. The results of the connectedness or diameter method are presented.

DIAMETER METHOD

	0	0	0	0	0	0	0	1	0	0	1
	2	1	5	4	6	7	9	0	8	3	1
0.600	.	XXX
0.570	XXXXX
0.510	XXXXXX	XXX
0.500	XXXXXX	XXX	XXX	.	.
0.480	XXXXXX	XXX	.	XXX	.	XXX	.	XXX	.	XXX	.
0.440	XXXXXXXXXX	.	XXX	.	XXX	.	XXX	.	XXX	.	.
0.420	XXXXXXXXXX	.	XXX	XXXXXX	.	XXX	XXXXXX	.	XXX	XXXXXX	.
0.410	XXXXXXXXXX	.	XXXXXXXXXX	.	XXXXXXXXXX	.	XXXXXXXXXX	.	XXXXXXXXXX	.	.
0.360	XXXXXXXXXX	XXXXXXXXXXXX	.	XXXXXXXXXXXX	.	XXXXXXXXXXXX	.	XXXXXXXXXXXX	.	XXXXXXXXXXXX	.
0.280	XXXXXXXXXXXXXXXXXXXXXXXXXXXX	.	XXXXXXXXXXXXXXXXXXXXXXXXXXXX	.	XXXXXXXXXXXXXXXXXXXXXXXXXXXX	.	XXXXXXXXXXXXXXXXXXXXXXXXXXXX	.	XXXXXXXXXXXXXXXXXXXXXXXXXXXX	.	.

POINTS IN A CLUSTER	CLUSTER STATISTIC
-----	-----
2 PTS.	1.901
1 5 (Information and Comprehension)	
3 PTS.	3.330
2 1 5 [Vocabulary and (information Comprehension)]	
2 PTS.	1.412
4 6 (Similarities and Sentences)	
2 PTS.	0.786
3 11 (Arithmetic and Block Design)	
2 PTS.	1.366
9 10 (Mazes and Geometric Design)	
5 PTS.	4.386
2 1 5 4 6	
3 PTS.	0.508
8 3 11 [Picture Completion and (Arithmetic-Block Design)]	
5 PTS.	1.769
9 10 8 3 11 [All performance tests]	
6 PTS.	1.459
7 9 10 8 3 11 [All Verbal tests]	

In this form of dendogram the objects are clustered from the top down. First the pairs 1 and 5 are formed at the lowest level. Next Vocabulary (02) is added to 01 and 05. Next 4 and 6 are combined, etc. Finally the spatial tests are combined with each other and the Verbal measures are combined at a correlation level 0.360.

The cluster statistic is designed to act as a Z score (Johnson,1967).

This page intentionally left blank

10

PARTITIONING

If researchers could see as well in n dimensions as they can in two, partitions of data points into a specified set of clusters could be done visually. In the two-dimensional example that follows (Figure 10.1), a good solution may be obtained by a simple inspection and placement of the data. For illustrative purposes, however, an objective solution is sought that is analogous to solutions obtainable for a larger number of data points measured on a larger number of variables or dimensions.

K-Means Iterative Clustering

The minimum distance method is as follows:

1. The data are initially assigned to one of a **prespecified number** of clusters. (This assignment can be made at random or by some other method).
2. The mean or centroid of each of the original clusters is determined.
3. Each data point is put into a new cluster that has the closest centroid to that data point.
4. The centroids or means of the new clusters are computed.
5. Alternate steps 3 and 4 until no data points change membership.

Table 10.1

Random and Iterative Assignments of Students					
Subjects	Reading	IQ	Random	1st	2nd
1	4	0	1	2	1
2	1	1	1	1	1
3	2	1	2	1	1
4	3	1	1	1	1
5	4	1	1	2	1
6	2	2	2	1	1
7	3	2	2	1	1
8	5	2	2	2	2
9	4	3	1	2	2
10	5	3	2	2	2
11	6	3	2	2	2
12	3	4	1	1	2
13	4	4	1	2	2
14	4	4	1	2	2
15	5	5	2	2	2

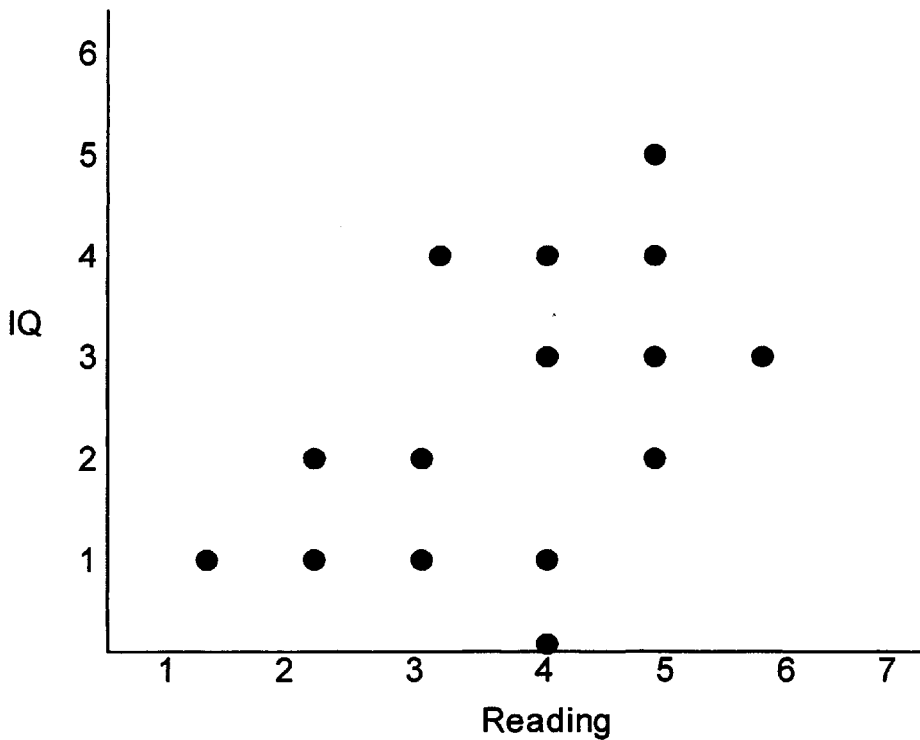


FIG. 10.1. Fifteen student's scores on measures of IQ and reading. Students are to be clustered into two groups.

Figure 10.1 illustrates the positions of the each of the students based on the coordinates of their Reading and IQ scores. Called **K-means** iterative clustering, classification into **two** groups is prespecified, that is, the number of groups is set at $K = 2$. The number of partitions is indicated prior to the analysis. In the example, a random assignment (coin flip) placed students into two groups. The centroids of these two groups are calculated by averaging the coordinates of all the subjects in each group. That is, the mean of IQ values for each initial group yields the IQ coordinate of the centroid for each group and the mean of reading values for each group yields the reading coordinate of the centroid for each group. Figure 10.2 shows the initial random assignment and the centroids of the two random groups. The data set is drawn from Table 10.1

The distances from all the data points to these two centroids were measured and points were reassigned based on the minimum distance to one of the two initial means. The process is repeated until no new assignment is possible. The resulting clusters and the new centroids are shown in Fig. 10.3.

The term "K-means iterative clustering" is understandable when we realize that k is the number of groups and thus the number of means that are prespecified. When dealing with many variables, over which a number of objects or subjects are to be classified, it is useful to employ a minimum variance criterion as a substitute for minimum distance. The "k-means" procedure is an attempt to minimize the variance within each cluster and consequently to maximize the variance between clusters. The sum of the squares of the distances of each point, within a cluster, from the centroid of that cluster, provides a within measure of variance. If, by relocating the data points, the sums of squares can be reduced, relocation takes place. Cluster solutions of this type have been called minimum variance partitions.

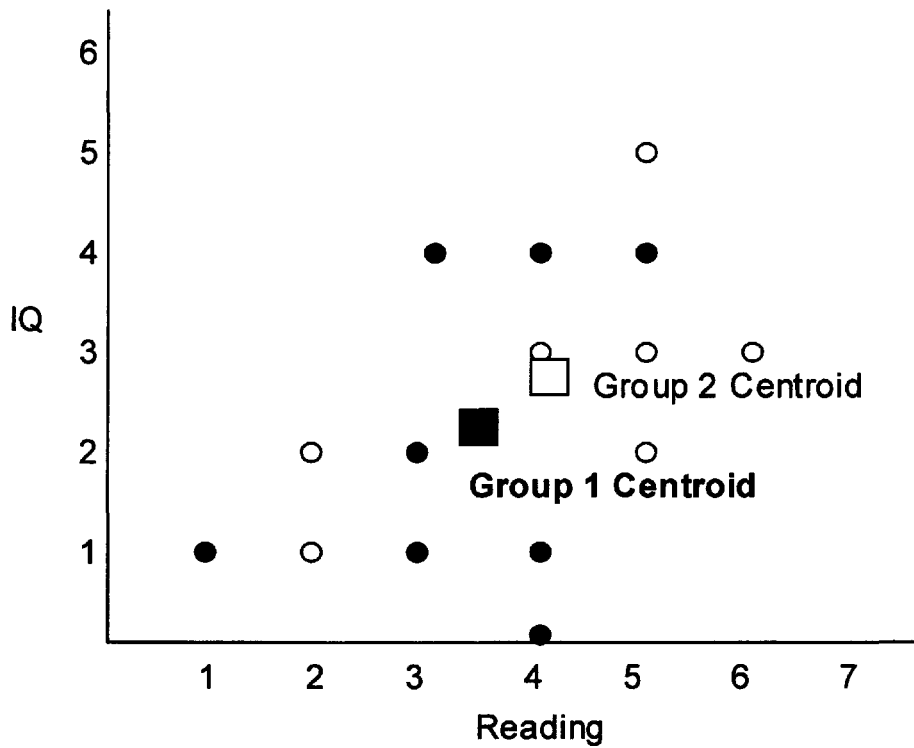


FIG. 10.2. Initial random assignment of students into one of two groups and the group means.

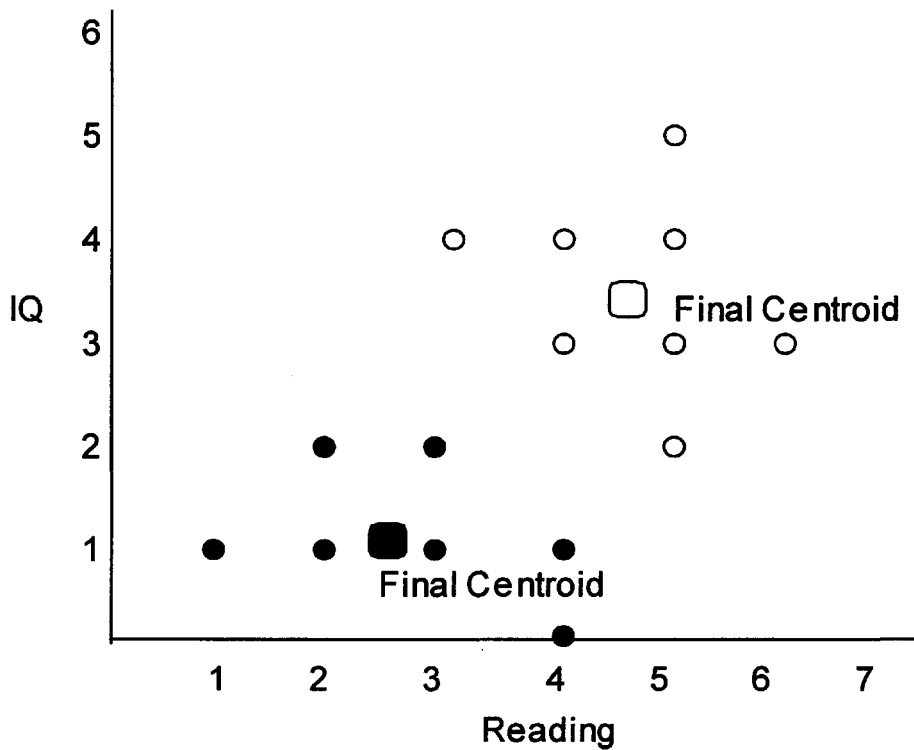


FIG. 10.3. Final assignment of students to two clusters.

Minimum variance partitions follow typical analysis of variance models. For example, consider a set of objects divided into two groups, say x and y . If an object belongs to x it will be called X ; if it belongs to y it will be called Y . The means of the two groups are:

$$\bar{X} = \sum X / N_x \quad \text{and} \quad \bar{Y} = \sum Y / N_y.$$

The Grand Mean (M) will be equal to

$$M = (\sum X + \sum Y) / (N_x + N_y).$$

Then, the total (T) sum of squares of all x and y from the grand mean (M) is equal to

$$T = \sum (X - M)^2 + \sum (Y - M)^2.$$

Within each group the sum of squares is

$$W_x = \sum (X - \bar{X})^2 \quad \text{and} \quad W_y = \sum (Y - \bar{Y})^2.$$

The sum of squares between the weighted means of each group equals

$$B = N_x(X - M)^2 + N_y(Y - M)^2.$$

From traditional analysis of variance we know that

$$T = W + B \quad \text{where} \quad W = W(\bar{x}) + W(\bar{y}).$$

This may be shown graphically as follows:

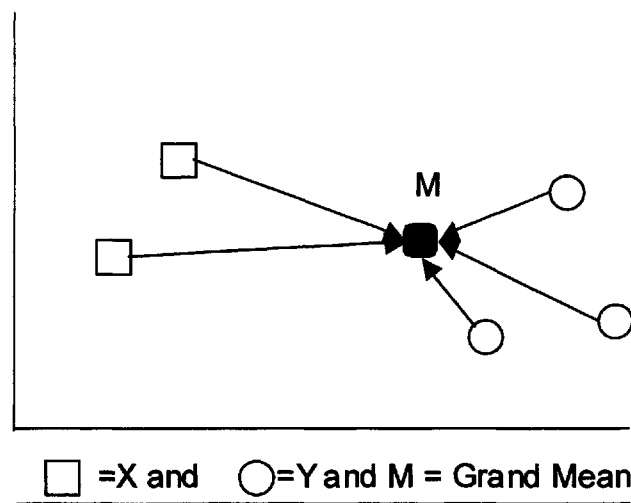


FIG. 10.4. Total sums of squares.

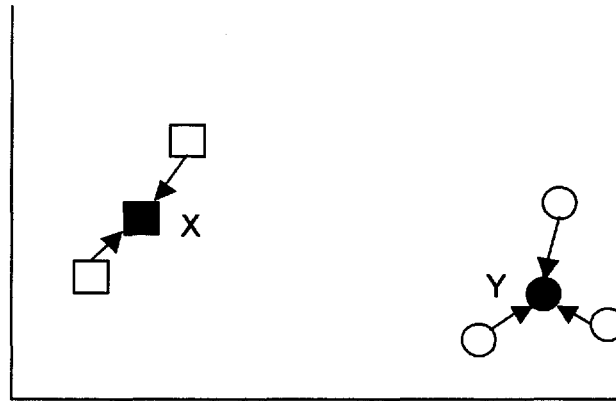


FIG. 10.5. The within sums of squares

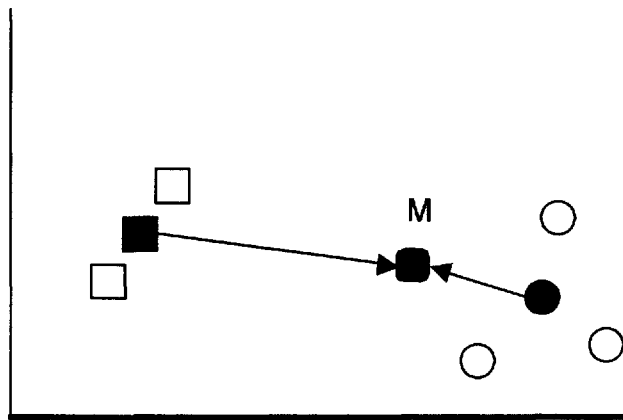


FIG. 10.6. The between sums of squares

In comparing various partitions of the objects it is reasonable to try to minimize $|W(i)|$. This is one of the options in most “k-means” partitioning programs

Application: Visual or Auditory Preference for Reading Instruction

Donovan (1977) attempted to classify children into three learning modality groups. These were Auditory Preference (AP), Visual Preference (VP), and No Preference (NP). She used a battery of ten diagnostic subtests from the *Illinois Test of Psycholinguistic Abilities* and the *Gates MacGinitie Readiness Skills Test*.

These tests were administered to 107 children in pre-reading programs. Donovan specified a clustering into three groups without specifying any basis for the clustering. A k-means iterative clustering using a computer program developed by McRae (1971) called MICKA used the minimization of the within cluster variance and distance criterion was Mahalanobis' d^2 (see p. 43).

The three clusters of children identified by the program were then compared with a clinical assessment of their modality preference. The clinical classification identified 36 pupils with Verbal Preference, 29 of whom were found in Cluster 2. Nineteen Auditory preference pupils were clinically classified, 14 of whom were found in Cluster 1. Clinical procedures identified 52 pupils as having No Specific Modality Preference. Twenty were found in Cluster 3.

	Clinical Assignment	K-Means Assignment	Percentage
Verbal Preference	36	29	0.81
Auditory Preference	19	14	0.74
No Preference	52	20	0.38

The results indicated that 78% of the children with an instructional modality preference (VP and AP) were identified through cluster analysis. The cluster means for the subtests are presented in Table 10.2.

Table 10.2
Cluster Means on Diagnostic Tests

Tests	Cluster 1	Cluster 2	Cluster 3
Visual Sequential Memory	37.0	38.4	28.9
Visual Discrimination	6.3	5.9	5.0
Visual -Motor Coordination	6.1	6.5	5.4
Word Recognition	6.4	6.3	5.8
Following Directions	4.3	4.1	3.7
Auditory Sequential Memory	44.5	32.5	33.7
Listening	4.8	4.1	4.2
Auditory Discrimination	6.1	5.6	5.1
Letter Recognition	7.2	7.1	6.6
Auditory Blending	5.3	5.2	4.8

Discussion

K-Means clustering is one of the few programs that allows for testing hypotheses. It allows the researcher to **prespecify** the number of groups of objects that will be delineated. If theory suggests there are only three types it can be tested. K-means clustering is part of the SAS packages of programs under PROC VARCLUS. For example:

```
PROC VARCLUS maxclusters = 3
outtree=new;
proc tree;
run;
```

11

HIERARCHICAL DIVISIVE

Successive Splitting

Hierarchical divisive methods, also known as disjoint cluster analysis, starts with all the objects in one cluster. This cluster is then divided into two clusters. Then one of these two clusters is subdivided resulting in three clusters. Next one of the three clusters is subdivided, and so on. The process continues until the N original objects are all separate clusters. Divisive clustering usually produces clusters in which all the members within a cluster are very similar. These are called monothetic classifications.

Dividing by Largest Variance

The Howard-Harris method (Blashfield & Oldenderfer, 1978) selects the one variable, in the set of variables, that has the largest variance. All subjects or objects with scores greater than the mean of this variable are placed in one group and all subjects with scores less than the mean are placed in the other group. This results in two initial clusters. The method uses a K-means iterative solution ($K = 2$) to determine the membership of the two clusters. Next, the one variable in the two beginning clusters that has the largest variance is selected and its mean is used to subdivide the high and low scores in that cluster. A K-means solution is applied to these two clusters to produce a three-cluster solution. As illustrated in Fig. 11.1, the process is repeated until some preset number of clusters is reached.

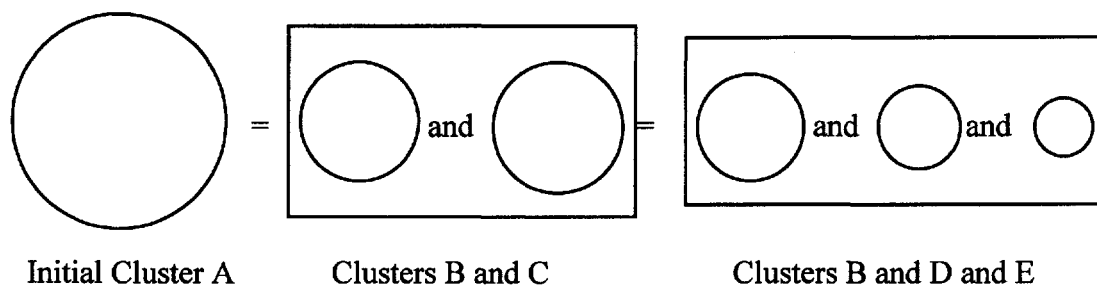


FIG. 11.1. Illustration of hierarchical divisive method of clustering.

The divisive clustering procedure is valuable because it can be applied to as many as 2000 subjects and as many as 20 variables. The method is part of the PC-MDS package of programs provided by BYU. The FORTRAN code for this method is given in *Scaling Methods* (Dunn-Rankin, 1983). SAS provides an extensive discussion of clustering methodology in its SAS documentation.

They state: "If you want to hierarchically cluster a data set that is too large to use with PROC CLUSTER directly, you can have PROC FASTCLUS produce, for example, 50 clusters, and let PROC CLUSTER analyze these 50 clusters instead of the entire data set." PROC FASTCLUS, unlike factor analysis, produces clusters that are not fuzzy. For example:

```
PROC FASCLUS Maxc = 3 nomiss;
Var V1-V20;
run;
```

(Recommended for large data sets)

Application: Grouping Ham Radios

J. Hills, one of the main author's students, had ten ham radio operators rank the similarity between six equipment manufacturers using paired comparisons. The radio manufacturers included Icom, Yaesu, Kenwood, Collins, Drake, and Ten Tech. The raw and standardized data are provided in Tables 11.1 and 11.2. Each profile has a mean of 3.5. The most frequent standard deviation is 1.71.

Table 11.1
Preference Profiles for Ham Radios

Radio	Subjects									
	1	2	3	4	5	6	7	8	9	10
Incom	4	1	3	3	2	4	1	3	2	1
Yaesu	3	3	2	3	1	3	6	4	3	4
Kenwood3	5	3	6	3	4	2	3	2	4	3
Collins	6	6	5	6	6	6	5	5	3	6
Drake	1	5	4	5	3	5	2	6	3	5
Ten Tech	2	3	1	1	5	1	4	1	6	2

Table 11.2
Standardized Preference Profiles for Ham Radios

Radio	Subjects									
	1	2	3	4	5	6	7	8	9	10
Incom	.29	-1.56	-.29	-.31	-.88	.29	-1.46	-.29	-1.19	-1.46
Yaesu	-.29	-.31	-.88	-.31	-1.46	-.29	1.46	.29	-.40	.29
Kenwood3	.88	-.31	1.46	-.31	.29	-.88	-.29	-.88	.40	-.29
Collins	1.46	1.56	.88	1.56	1.46	1.46	.88	.88	-.40	1.46
Drake	-1.46	.93	.29	.93	-.29	.88	-.88	1.46	-.40	.88
Ten Tech	-.88	-.31	-1.46	-1.56	.88	-1.46	.29	-1.46	1.99	-.88

Howard-Harris divisive clustering was utilized to group the manufacturers

The two-group and three-group solutions are presented below

1	2	1	2	3
Icom	Collins	Icom	Collins	Ten Tech
Yaesu	Drake	Yaesu	Drake	
Kenwood		Kenwood		
Ten Tech				

Hills was able to make sense of the three-group solution because Group 1 were all Japanese manufacturers, Collins and Drake were similar U.S. companies, and Ten Tech was a new company.

Number of Clusters

Although an appraisal of the cluster solution (number of clusters) in Wards' method can be made based on the size of the sums of squares, two other methods can be employed. One is to split the original data and cluster both sets. If the clustering is similar in both cases, it can be concluded that a reliable solution has been obtained. The clustering can also be replicated.

Another method is to use an outside criterion consisting of a set of predictor variables. This method has the following steps: (1) Select a reasonable range of groups, and (2) Sequentially predict group membership using discriminant function analysis (multiple regression used to predict membership into varying numbers of clusters). The chi-square value for each cluster solution divided by the resulting degrees of freedom can be used as a selection index (i.e., where chi square/df is maximum).

Graphing the Clusters

All clustering should be presented pictorially in order for the information to be easily understandable. The basic clustering dendrogram can be thought of as a mobile. In this mobile, each horizontal bar is free to rotate around the connecting vertical line. Figure 11.2 consists of words beginning with the letter h. In Fig.11.3, a rearrangement has been made so as to afford a better pictorial contrast between short and long words. Chambers and Kleiner (1980) suggested different ways of displaying the results of clustering. The dendrogram can be altered by shortening the end lines or by making a straight line connection to the nodes (points of departure from a continuous line; see Fig. 11.4).

It will be seen in Part IV, Multidimensional Methods, that clustering can be pictorially superimposed on a dimensional solution by drawing circles around clusters of plotted objects.

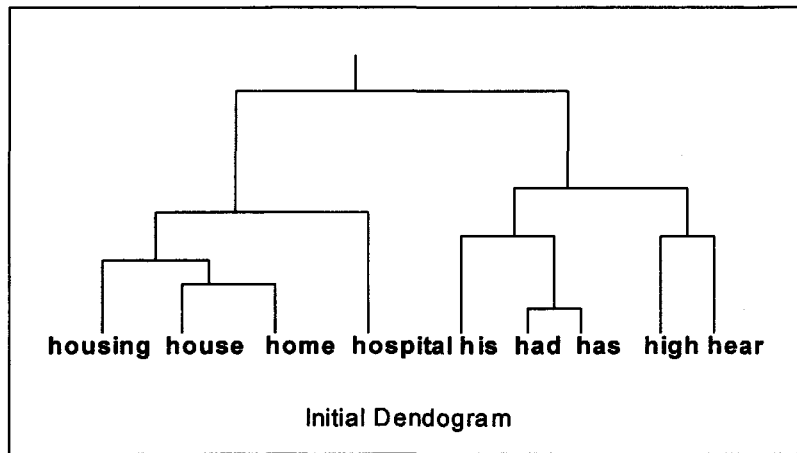


FIG. 11.2. Regular Dendrogram.

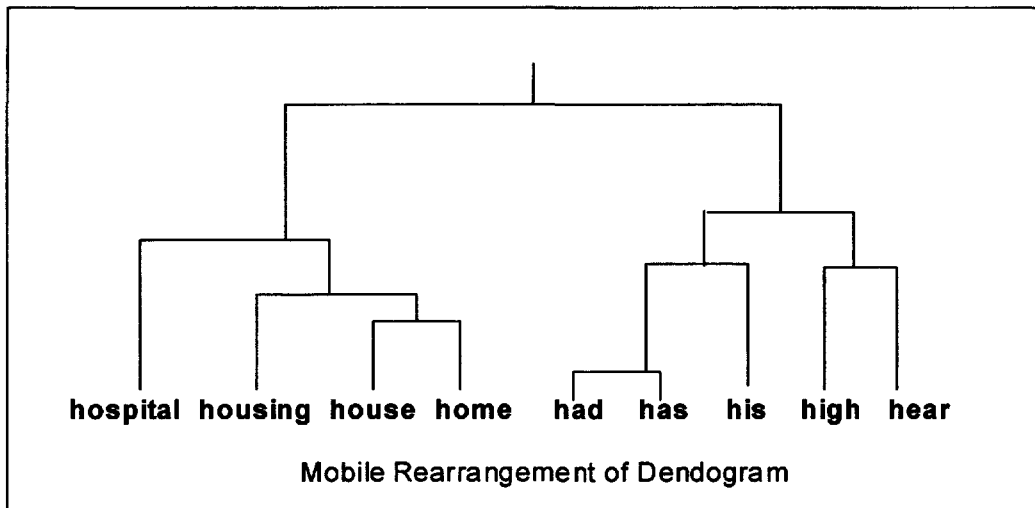


FIG. 11.3. Dendrogram rearranged to enhance the difference between clusters.

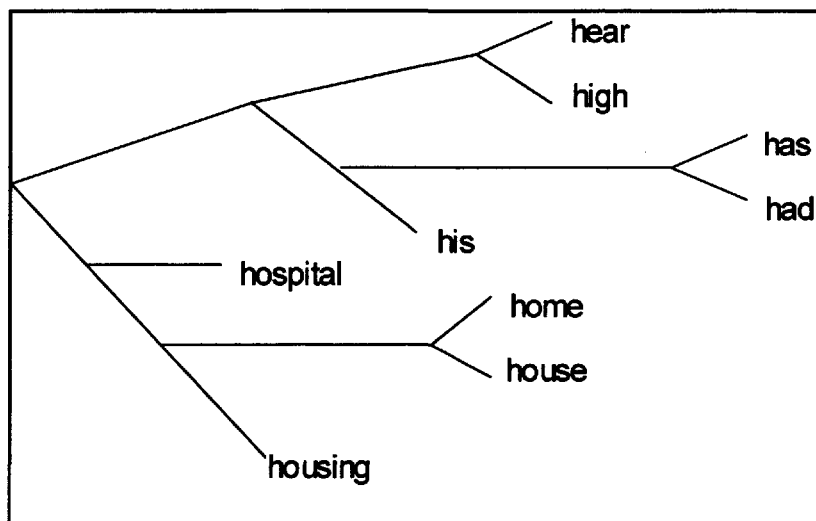


FIG. 11.4. A branching Dendrogram.

PART IV

MULTIDIMENSIONAL METHODS

Part IV presents four useful multidimensional scaling methodologies. The technique of **factor analysis**, traditionally developed and utilized with tests of ability and achievement, has also been applied extensively to the reduction of matrices of correlations. Factor analysis however, contains restrictive assumptions of linearity between variables and homogeneity of variance among the variables. **Singular value decomposition** (SVD) can be applied to any rectangular matrix. It is used in **multidimensional preference analysis** (MDPREF) The simpler assumptions underlying **multidimensional scaling** allow the methodology to be utilized in a different and generally more concise description of a data matrix. **Preference mapping** and **individual differences scaling** are extensions of factor and multidimensional scaling analyses that provide insights into how individuals differ with regard to the same psychological objects.

This page intentionally left blank

12

FACTOR ANALYSIS

Representation of the Correlation Matrix

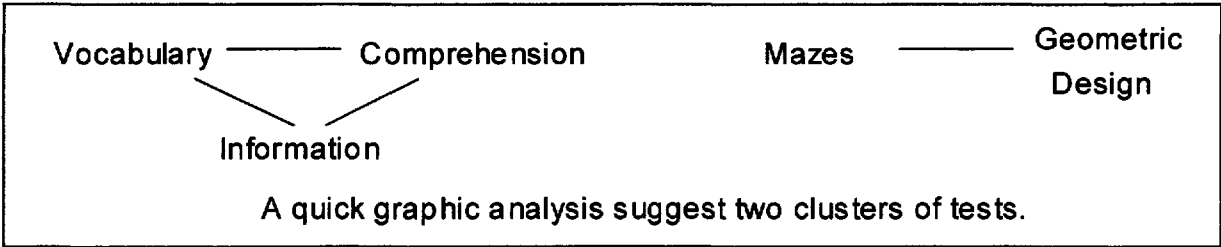
Factor Analysis attempts to simplify a large body of data by identifying or discovering redundancy in the data. The factors are smaller representations that are derived from a larger matrix of correlations. In psychological research intercorrelations among a set of tests, such as the subtests of Wechsler Intelligence Scale for Children (WISC), reveal groups of subtests that are interrelated. In Table 12.1 are intercorrelations among five subtests of WISC.

Table 12.1

Intercorrelations Among Five Subtests of the WISC

	Information	Vocabulary	Comprehension	Mazes	Geometric D.
Information	1.00	.60	.60	.37	.40
Vocabulary	.60	1.00	.57	.35	.35
Comprehension	.60	.57	1.00	.28	.30
Mazes	.37	.35	.28	1.00	.48
Geometric D.	.40	.35	.30	.48	1.00

Are the relationships among the three Verbal tests redundant? That is, can one predict a person's Comprehension score by her or his score on Vocabulary? Can scores on Mazes predict scores on Geometric Design? The correlations suggest that some prediction is possible and suggests that the matrix contains overlapping information or is redundant to some degree. A graphic analysis suggests two groups of subtests, one verbal and one visual.



If a matrix is redundant, its **true rank** (the number of **independent** rows or columns) will be less than its original rank. The original rank is the number of variables or subtests. In our example, the original rank is five (5) since there are five subtests. The total variance of the matrix is also five. This can be found by noting that each variable, when standardized as Z scores, initially contributes 1 unit to the total variance. The values of the diagonal elements in the initial matrix are all 1.00 and sum to 5.00.

The question to be answered by Factor Analysis is: What is the **approximate true rank** of the matrix? Approximate because the true rank can only be estimated. This is because errors of measurement in the social sciences do not allow the researcher to find truly dependent variables, that is, perfect correlations. If one estimates that there are two factors that represent the five subtests in our example, the suggestion is that the approximate rank of the matrix is two (2).

The elements of a factor are the correlations of each test with the factor, called factor loadings. The factor matrix, also known as the factor pattern, is represented by F. If one can estimate the elements of each factor, then by multiplying the factor matrix by its transpose ($F \bullet F'$) an effort is made to recreate the original correlation matrix (R). A comparison of the new (R^*) matrix with the original (R) matrix indicates how successful the estimate has been.

The transpose of F exchanges the rows and columns as illustrated below. Matrix multiplication is demonstrated in the R^* matrix. Note F is a 3 rows by 2 columns matrix. F' is 2 by 3 and this results in a 3 by 3, R^* matrix.

F	
I	II
a	d
b	e
c	f

X

F'	I	a	b	c
	II	d	e	f

=

R*		
$a^2 + d^2$	$ab + de$	$ac + df$
$ba + de$	$b^2 + e^2$	$bc + ef$
$ca + fd$	$cb + fe$	$C^2 + f^2$

Trial and Error

Based on the **Graphic Analysis** (p. 115), it is assumed that there are two factors. How can one estimate the values to be inserted in the factor matrix? One way is trial and error. The intercorrelations between the first three variables are approximately .60. In matrix multiplication with two factors (FF'), .60 can be determined by the sum of two products ($.60 = ab + de$). What values will best satisfy this equation? That is, what should the Factor matrix look like? First a number of trials are made to search for acceptable elements of the F matrix as shown below.

Two estimates, .7 and .3 appear to be good choices for the first **three** tests. This selection is based on approximating the other elements in the R matrix away from the diagonal. $(.7)(.7) + (.3)(.3) = .58$, for example, which is close to .60.

The same question can be asked for the intercorrelations between the last two variables (Mazes and Geometric Design) which is .48. Here we select .3 and .6 whose summed products $(.3)(.3) + (.6)(.6) = .45$, which is close to the recorded value of .48.

Now we put our best estimates into a factor matrix as shown below in Table 12.2. We designate the factors by the letters **a** and **b**.

Trial Factor Matrices											
a	b	a	b	a	b	a	b	a	b	a	b
1.0	0	.9	.1	.8	.2	.7	.3	.6	.4	.7	.3
1.0	0	.9	.1	.8	.2	.7	.3	.6	.4	.7	.3
1.0	0	.9	.1	.8	.2	.7	.3	.6	.4	.7	.3
0	1.0	.1	.9	.2	.8	.3	.7	.4	.6	.3	.6
0	1.0	.1	.9	.2	.8	.3	.7	.4	.6	.3	.6

Table 12.2

Final Trial Factor Matrix, F

	Factor a	Factor b
Information	.70	.30
Vocabulary	.70	.30
Comprehension	.70	.30
Mazes	.30	.60
Geometric D.	.30	.60

In multiplying the Factor matrix (**F**) by its transpose (**F'**), the row and columns are interchanged and the product is found as follows:

Table 12.3
Matrix Multiplication of the Factor Matrix by
its Transpose ($FF' = R^*$)

		info	voc	comp	Maze	Geo.
	a	.7	.7	.7	.3	.3
a	b	.3	.3	.3	.6	.6
.7	.3	.58	.58	.58	.39	.39
.7	.3		.58	.58	.39	.39
.7	.3			.58	.39	.39
.3	.6				.45	.45
.3	.6					.45

The **off diagonal** elements of the recreated R^* matrix are subtracted from their corresponding original correlations and the average absolute difference is calculated. The average difference equals .039. This difference is small and fairly evenly distributed. The factor matrix (**F**) is a reasonable representation of the original correlations of the **R** matrix.

The original diagonal elements (1.00) are not closely approximated by matrix multiplication. The values in the diagonal (see Table 12.3) are closer to the Squared Multiple Correlations ($SMC = r^2$) of each variable with all the others. When SMC are placed in the diagonal of the correlation matrix, then the Factor Analysis will analyze the **common variance**, rather than the **total variance**, in the data.

Predicting one variable from a number of other variables is a problem in multiple regression analysis. The r^2 that results from this analysis is a measure of the accountable variance in prediction (Pedhazuer and Schmelkin, 1991).

Test Score Assumptions

The factor loadings can derive meaning from basic assumptions surrounding test scores. First, a test score is assumed to be the sum of a number of components. This is called the additive assumption. A particular intelligence test score may, for example, consist of the sum of separate contributions made by **verbal** and **numerical** ability, plus a component **specific** to the individual items plus a component due to **error**. Symbolically, after standardization, this assumption can be written as follows:

$$Z_{\text{test}} = Z_v + Z_n + Z_s + Z_e$$

Z_v and Z_n , the verbal and numerical components, are called common factors because they are dimensional elements found in both subtests. The extent to which any two tests are related depends on the extent to which their common factor components are similar. This does not include the specific or error components. The assumption of additivity of the components of a total score can be more formally stated as follows:

$$Z_t = aZ_{x_1} + bZ_{x_2} + cZ_{x_3} + \dots + qZ_{x_k} + sZ_{x_k} + tZ_e$$

common components specific error

where a, b, \dots, q are weights assigned to each component.

The equation is expressed in standard Z form. This equation states that a total test score is a weighted summation of **common factor** scores plus a **specific component** plus an **error component** (Guilford, 1954, p. 476). The assumption can also be applied to the variance of the scores:

$$s_t^2 = s_a^2 + s_b^2 + s_c^2 + \dots + s_q^2 + s_s^2 + s_e^2$$

Total Variance	=	Common Variance	+	Unique Variance.
----------------	---	-----------------	---	------------------

The total variance of the scores on a test or variable may also be subdivided into three general types - common, specific, and error variance. **Common variance** is that portion of the total variance that correlates with other variables. **Unique variance** is the sum of specific variance and error variance. **Specific variance** is that portion of the total variance which does not correlate with any other variable. **Error variance** is chance or random variance, due to errors of sampling, measurement, . . . and "the host of other influences which may contribute unreliabilities," (Fruchter, 1954, p. 45).

Each component of the variance equation can be expressed as a proportion of the total variance. For convenience, simpler terms are substituted for the proportions. The variance equation then becomes

$$1.0 = a_x^2 + b_x^2 + c_x^2 + \dots + q_x^2 + s_x^2 + e^2.$$

Accountable Variance

h_x^2 is called "**communality**." Communality is defined as the **sum** of the proportions of common factor variance in the test score. **Uniqueness** (u_x^2) is the portion of the total variance that is not shared in common with any other variable. And **specificity** (s_x^2) is the proportion of specific variance in a test or variable.

$$u_x^2 = s_x^2 + e^2 = 1 - h_x^2$$

Because u_x^2 can be obtained as $(1 - h_x^2)$, u_x^2 is not computed directly in factor analysis. In other words, the specific and error components are not estimated. Factor analysis relies on the factor loadings on common factors to approximate the R matrix. The factor coefficients or loadings exist as correlations (i.e., square roots of variance components). They are **correlations of the particular test with the factor**. In our example, Vocabulary correlates (.70) most highly with Factor a. Mazes (.60) correlates more highly with Factor b. The loadings are squared, in order to provide estimates of variance (any r^2 is the proportion of variance accounted for). For the example given in Table 12.4, the sum of the squares of the factor loadings for each **row** is calculated to obtain h^2 or the communality. This is an estimate of how well the factors account for the variance in each test.

Summing the squares of the factor loadings for each **column**, the sum is **1.65** for the first factor and **.99** for the second. These are sometimes called the eigenvalues (λ)* in a statistical analysis. The total of **2.64** is equal to the sum of the diagonal elements of the new R^* matrix, also known as the estimated total communality. This total is subtracted from **5.0**, the total variance, which leaves **2.36** left over. Therefore, these two factors ($2.64/5.0$) account for **53%** of the total variance.

Table 12.4

Accounting for Variance Using Factor Loadings

Tests	Factor Loadings		Factor Variance		Communality	Uniqueness	Total
	a	b	a^2	b^2	h^2 (a^2+b^2)	u^2 $1-h^2$	
Info.	.7	.3	.49	.09	.58	.42	1.0
Vocab.	.7	.3	.49	.09	.58	.42	1.0
Comp.	.7	.3	.49	.09	.58	.42	1.0
Maze	.3	.6	.09	.36	.45	.55	1.0
Geo. D.	.3	.6	.09	.36	.45	.55	1.0
Sums λ			1.65	.99	2.64	2.36	5.0
%			.33	.20	.53	.47	1.00

*Eigenvalues are the solutions to the characteristic equation used in some formal solutions of factor analysis such as principal components analysis. Eigen comes from the German word for root, that is, root of the equation.

Principal Component Analysis (PCA)

There are more objective and sophisticated ways of determining the best factors rather than using **trial and error**. Just as in estimating the best fitting line in regression analysis, similar procedures can be used to estimate the best factors. Factor solutions take advantage of matrix algebra. Factor analysis (such as principal components) is used to reduce the array of data in a correlation matrix to a smaller factor matrix. A principal concept in this theory is the **rank** of a matrix. Fortunately, given any matrix and using computers, there are standard methods for determining its approximate rank (Harman, 1967, Rummel, 1970). By using SAS for the Personal Computer, a template for **principal component analysis (PCA)** starting with a correlation matrix is as follows:

```
/*Principal Components (PC) Analysis*/
data wisc (type = corr);
  _type_ = 'corr';
  Input _name_ $ inform vocab compre mazes geodsn;
  datalines;
inform 1.00 0.60 0.60 0.37 0.40
vocab 0.60 1.00 0.57 0.35 0.35
compre 0.60 0.57 1.00 0.28 0.30
mazes 0.37 0.35 0.28 1.00 0.48
geodsn 0.40 0.35 0.30 0.48 1.00
;
proc factor method=principal
priors=one scree nfact=2 ,
rotate=varimax reorder;
run;
```

Title

Name of the SAS data set

Need if data is an r matrix

datalines

Each part ends with ;

Asks for PC analysis

Ones in Diagonal, a scree test, 2 factors,

Rotate using varimax, Order factors

The results of **PROC Factor** for our WISC correlations are as follows:

The FACTOR Procedure
Initial Factor Method: Principal Components

Prior Communality Estimates: ONE

*(These variance **estimates** are placed in the diagonal. Their sum = 5)*

Eigenvalues of the Reduced Correlation Matrix: Total = 5 Average = 1

	Eigenvalue	Difference	Proportion	Cumulative
1	2.7389	1.8054	0.5478	0.5478
2	0.9335	0.4119	0.1867	0.7345
3)	0.5216	0.0973	0.1043	0.8388
4)	0.4242	0.0426	0.0848	0.9237
5)	0.3815		0.0763	1.0000

(The calculated eigenvalues are divided by the estimate (5.0) to get the proportion.)

For this example, two factors will be retained by the NFACTOR criterion.

Eigenvectors (V)

	1	2
inform	0.4986	-0.2384
vocab	0.4809	-0.2894
compre	0.4606	-0.4286
mazes	0.3868	0.5999
geodsn	0.3977	0.5618

Factor Pattern (VL⁻⁵)

	Factor 1	Factor 2
inform	0.8252	-0.2303
vocab	0.7959	-0.2796
compre	0.7623	-0.4141
geodsn	0.6582	0.5429
mazes	0.6402	0.5796

(The initial factor loadings are presented)

Variance Explained by Each Factor

Factor 1	Factor 2
2.7389	0.9335

(These are eigenvalues, λ , the sum of squares of the factor loadings)

Final Communality Estimates: Total = 3.6725

inform	vocab	compre	mazes	geodsn
0.7341	0.7117	0.7526	0.7458	0.7280

(These are h^2 the sum of the squares for each row)

Rotation Method: Varimax

(Kaiser's maximum variance criterion is used rotate uncorrelated factors)

Orthogonal Transformation Matrix

	1	2
1)	0.8131	0.5820
2)	-0.5820	0.8131

(The varimax criterion indicates a rotation of 36° 44'. Cos = .801, Sin = .582)

Rotated Factor Pattern

	Factor 1	Factor 2
Com	0.8609	0.1069
inf	0.8100	0.2359
voc	0.8015	0.2930
geo	0.1831	0.8439
maz	0.2192	0.8246

(The original factor pattern was post multiplied by the transformation matrix)

Variance Explained by Each Factor

Factor 1	Factor 2
2.1272	1.5452

Factor Rotation

In the statistical analysis, a rotation is performed on the initial factor pattern. The orthogonal transformation matrix is of the form:

$$\begin{vmatrix} \cos Q & \sin Q \\ -\sin Q & \cos Q \end{vmatrix}$$

where Θ is 36.8 degrees. In Figure 12.1 the original WISC factor loadings are plotted using the horizontal and vertical axes for Factors 1 and 2. A 37 degree rotation sharpens the analysis by reading the new loadings (bold type) off the rotated axes.

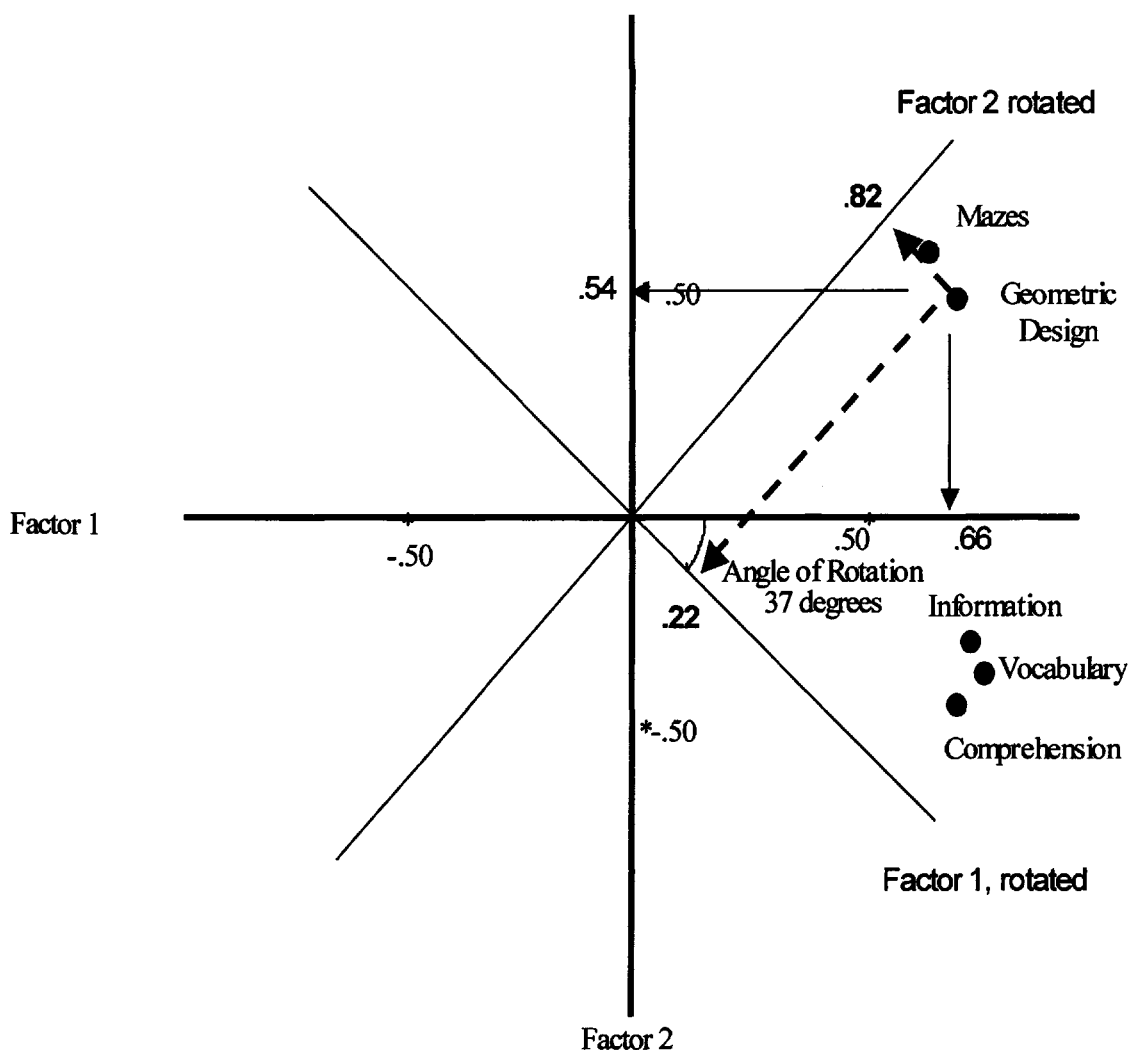


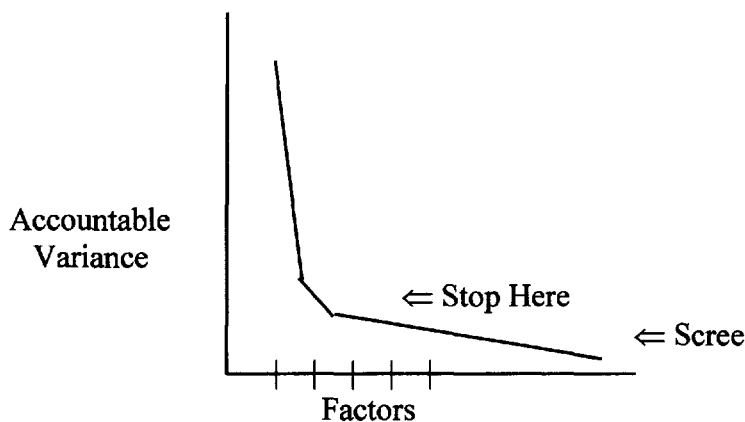
FIG. 12.1. Illustration of a 37° rotation.

For example, if one reads the approximate coordinates of geodsn (Geometric Design) on the original axes, Factors 1 and 2 are .658 and .542. Reading these same values from the rotated axes yields .22 and .82. The rotated values are approximated by the trial and error method. No rotation was performed on the experimental solution indicating that trial and error values have utilized some theoretical justification in the initial solution.

Specific Problems Associated With Factor Analysis

1. What value should be placed in the diagonal of the correlation matrix? Placing 1 in each diagonal entry indicates that the experimenter is interested in all of the variance in the data, specific and error as well as common. Placing some other value in the diagonal indicates that an estimate of the communality is available and is meaningful. The largest row correlation (without regard to sign) or squared multiple correlations (SMC) have been suggested as the best values to place in the diagonal. Harman (1967) has a good discussion of these issues.

2. How many factors should be extracted? Kaiser (1958) suggested that if the sum of the squares of each set of factor loadings is less than (1) little is to be gained by extracting further factors. He reasoned that initially each variable contributed a value of 1.0 and questioned what factor could contribute less than a variable. Cattell (1962) suggested that the decision on the number of factors be based on what is called the Scree Test. The Scree Test consists of plotting the factors in equal intervals on the x axis against the variance accounted for by each factor.



Cattell (1966) represented a line plot of the variances accounted for by the factors as similar to the erosion from a tall cliff. First there is a precipitous drop from the first one or two factors down to the rubble accumulating at the base of the cliff or the scree. The first one or two factors usually account for the most variance and then the other factors account for less and less variance until a straight line will fit the remaining variances. The point at which there is a break between the downward fall of the variance and the straight line representing the scree is usually the cutting point for the number of factors.

Reliable Component Analysis (Cliff & Caruso, 1998) provide a useful method for determining the number of factors or components that should be retained.

3. What kind of rotation should be performed? The initial solution is always based on the assumption of uncorrelated factors. This is known as the “orthogonality” restriction imposed on PCA analysis. Usually the general factor solution that emerges before rotation is not as interpretable as a solution that can be obtained by rotating the axes in order to contrast the factor loadings more effectively. The most popular rotational procedure is Kaisers’ orthogonal varimax rotation. This process attempts to rotate the axes so as to maximize the variance accounted for by each factor. This is done by minimizing the sum of the rectangular areas encompassed by the coordinates of the points and the axes from which they are drawn. Non-orthogonal or oblique rotations allow for the factors to be correlated. If one is interested in higher order factors, then an oblique rotation allows a matrix of factor correlations to be analyzed.

This page intentionally left blank

13

MAPPING INDIVIDUAL PREFERENCE

In determining the preference for almost any entity (Pepsi, grape jelly, University of North Florida, Tiger Woods, high school, etc.), a directional identification can be made. That is, given two or more psychological objects, subjects usually prefer one of the elements of a set. Individual differences in preference are of interest to the behavioral scientist because attitude by treatment interactions have not been fully explored. Different people may react differently to the same stimulus. Some people prefer spinach, others dislike it. Some children prefer teacher approval as a reward; others prefer freedom, competitive success, peer approval, or consumable rewards. The methods previously discussed have looked at psychological objects from the view of the average subject. A multidimensional mapping of objects (like a unidimensional scale) has also been represented as the average respondent's judgment or preference between pairs of objects. It is important, however, to look at the specific individual's preference. That is, to study individual interaction between attitude and treatment

Singular Value Decomposition

Before multidimensional preference analysis is examined, a brief introduction to matrix reduction (estimating the approximate rank) using Singular Value Decomposition is needed. Carroll and Chang (1970) introduced the use of this process in preference analysis.

Many multidimensional analyses are based on a **matrix** theorem of Ekart and Young (1936) Their theorem indicates how the largest or most important eigenvalues of a correlation matrix can be used as a quantitative measure of the approximate rank of a matrix. The **singular values** (the solution) are the square roots of the eigenvalues given in descending order of importance. They end up as values in the **diagonal** matrix. Just as any number like 42 can be factored in its primes (2)(3)(7), The theorem indicates that any rectangular data matrix can be put in **diagonal form** with the aid of Singular Value Decomposition (SVD), which means:

$$\begin{array}{ccccccc}
 \boxed{D_{LM}} & = & \boxed{R_{LL}} & \times & \boxed{A_{LM}} & \times & \boxed{S_{MM}} \\
 \text{Initial Matrix} & & \text{A Singular} & & \text{The} & & \text{A Singular} \\
 & & \text{Matrix} & & \text{Diagonal} & & \text{Matrix} \\
 & & & & \text{Matrix} & &
 \end{array}$$

$$D_{LM} = R_{LL} A_{LM} S_{MM}$$

Where R_{LL} and S_{MM} are different singular matrices and A_{LM} is diagonal matrix whose entries are the singular values. The **rank** of the product is equal to the rank of the diagonal matrix when its extremely small values are discounted. The research decides how small these values should be. It could be a ratio of 1/100 of the total variance or less. For example

$$D = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} = \begin{bmatrix} -.405 & -.915 \\ -.915 & .405 \end{bmatrix} \times \begin{bmatrix} 5.47 & .0 \\ .0 & .366 \end{bmatrix} \times \begin{bmatrix} -.576 & -.817 \\ .817 & -.576 \end{bmatrix}$$

Initial

Singular

Diagonal
Solution

Singular

The reader can verify that the product of the three matrices on the right equals the original matrix. In the middle is the diagonal matrix. Thus a diagonal matrix has been found whose diagonal elements are the square roots of the singular values. The singular values are ordered on the diagonal and 0.366 is very small compared to 5.47. In practice, very small values indicate redundancy in the data and suggest a matrix of lower rank can be utilized to explain the data.

Carroll and Chang's Multidimensional Vector Model

The vector model of preference mapping (Carroll, 1972) is analogous to scoring a subject's preference, on a unidimensional scale, in the multidimensional space of the objects. The process usually starts with a two or three-dimensional configuration of objects whose interpoint distances (the distances between the objects in t dimensions) have been derived from judgments of their similarity. The subject's preference direction is then included in that configuration.

Suppose, for example, the similarities among four desserts: chocolate cake, chocolate ice cream, pound cake and vanilla ice cream were judged by a group of children and the resulting configuration of the four desserts is as shown in Fig. 13.1.

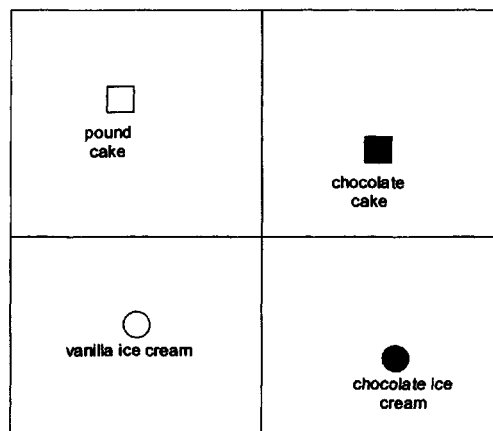


FIG. 13.1. Configuration of desserts.

It is easy to label the dimensions as Chocolate versus Non-chocolate and Cake versus Ice cream. Next, suppose a child was asked to rank order her or his preference for the four desserts and the results were as follows:

Child A— (1) Chocolate Cake, (2) Chocolate Ice Cream, (3) Pound Cake, (4) Vanilla Ice Cream

Surprisingly, the preference direction, for this subject, can be estimated by the constraints imposed by the initial configuration of desserts on the child's rank order of preference.

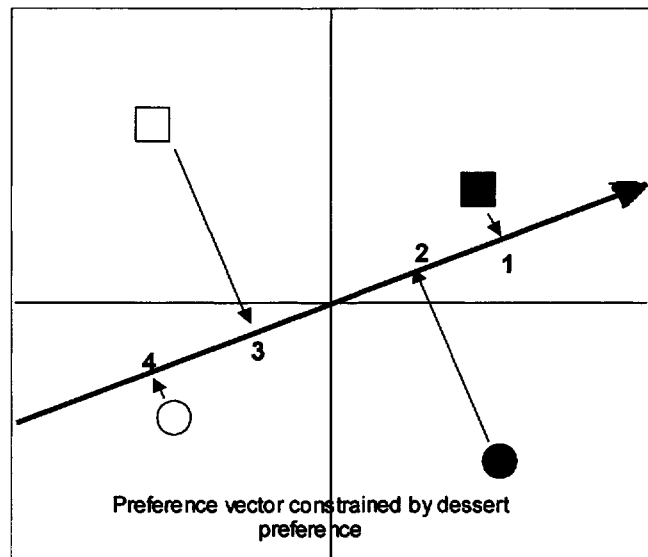


FIG. 13.2. Perpendicular projection preserves rank order. The ranks determine the vectors direction

As shown in Fig. 13.2 the perpendicular projection of the stimuli preserves the respondent's preference values (rank order). The direction of the vector is of particular interest because it reveals individual differences in preference with regard to the dimensions represented by the desserts. A large number of different vectors may be accommodated in a two dimensional space. When there are several objects and their configuration has been well defined (as by the children, in this case), the direction of each subject's preference vector is uniquely determined. Generally, the stimulus configuration and preference vectors (direction) are analyzed from the same data. Preference mapping, using the vector model, assumes that the stimulus points (objects) are projected onto individual subject's vectors.

It is possible to start the analysis with a predetermined configuration of objects. The case in which the object configuration is determined in advance of the preference mapping has been called external analysis.

If the respondent's vectors are of unit length, the projection of an object vector onto the respondent's vector can be obtained from the scalar product. If, for example, Y_i is the unit length vector for student C , and X_j is the vector for object j , then S_{ij} , the theoretical scale value of X_j for Y_i , is:

$$S_{ij} = Y_i X_j \cos \Theta$$

As shown in Fig. 13.3, Y_i is the vector for Student C and X_j is the object. The object is projected onto the student's vector.

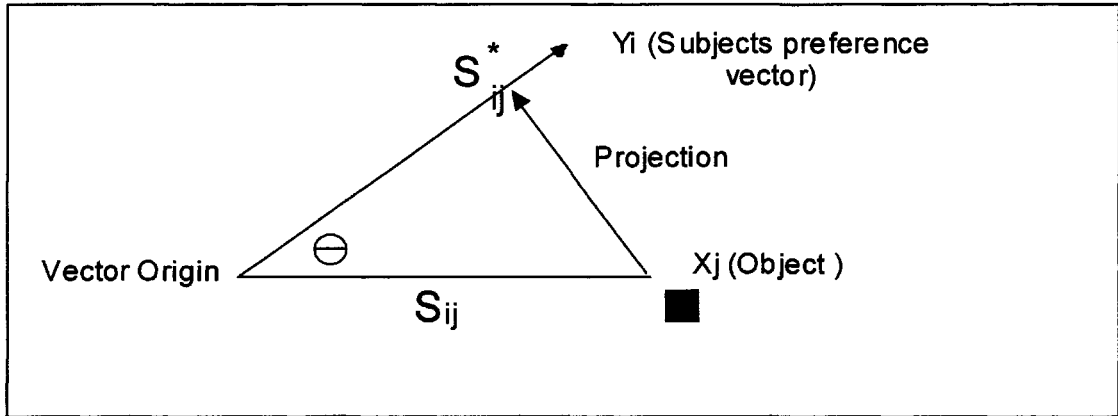


FIG. 13.3. Illustration of object projection creating a theoretical scale value.

Therefore, given a respondent's preference for a set of stimulus objects such as X_1, X_2, \dots, X_J with a preference rank of 1, 2, ... J, the problem is to find the vector that best fits the stimulus projections. The object projection (scalar product) onto any specific vector is proportional to the distance between the vector's end point and the stimulus point. The question can be resolved by determining that specific point where the vector terminates on a circle enclosing the space of the objects. The circle's origin is the centroid of the objects. Mathematically the problem of preference mapping is to find the slope of each respondent's preference vector. The problem becomes one of minimizing the sum of the differences between the preference values (ranks) S_{ij} and the distances from the objects to the vector terminus on the circle that are proportional to the projections S^*_{ij} on a vector through the origin; that is,

$$\sum (S^*_{ij} - S_{ij})^2 \text{ is to be minimum, which is the typical problem in linear prediction.}$$

MDPREF

MDPREF is a factor analysis, using Singular Value Decomposition that not only produces the factor structure of the objects but processes the subjects as well. Carroll and Chang (1968) wrote a program that performs a linear factor analysis on the stimuli and fits preference vectors to the object configuration. It has particular usefulness in perception and attitude measurement where rank profiles of preference are available. Their program plots the individual's vector direction within the object space. MDPREF has the unique advantage of being able to handle a large number of participants if the number of objects is relatively small.

MDPREF is what is known as a “vector model”. This means that the objective of the MDPREF analysis is to identify a map displaying the participant’s preference vectors. The model assumes that preference is greatest near the end of his or her vector. The stimuli or object points are plotted in the space. To form the vectors visually, lines are drawn from the origin of the plot to each participant’s point on a circle.

Each **stimulus** point projects (at a 90 degree angle) onto each subject’s vector. This projection attempts to preserve the preference for the stimuli. More technically, MDPREF is a metric model based on a principal components analysis (Ekart-Young decomposition, SVD). In this analysis, a data matrix of *i* subjects by *j* stimuli is decomposed into two smaller matrices, each of which approximates the original data matrix using least squares. The first of these resulting matrices is a principal component score matrix of size (*i* x *t*), where *t* is the number of dimensions. This matrix depicts the *j* subjects in the *t* dimensions. The second matrix is called the principal component matrix and is of size (*t* by *j*). This matrix depicts the *j* stimuli in the *t* principal component dimensions.

CD-ROM Example Using MDPREF

An example related to the four desserts: Chocolate Cake, Pound Cake, Vanilla Ice-Cream, and Chocolate Ice-Cream is presented. Seven adults voted their preference for one of each of the six possible pairs. In this data the higher the value the more it is preferred. The program, **MDPREF**, is found on the CD-ROM. First a configuration file is created:

mdpref.cfg

```
This is the Second Trial for MDPREF on Four Desserts
  7  4  2  2  2
mdpref.dat
mdpref.out
```

Title
#persons = 7, #items = 4, #factors = 2, #factors
Plotted = 2, (standardization, 1= subtract mean
from each score, 2 = make Z score).

mdpref.dat

```
2 3 3 2
2 1 4 3
2 1 4 3
1 4 2 3
3 2 2 3
3 1 2 4
3 2 2 4
```

In this data, note that the first subject least preferred items 1 and 4 and most preferred 2 and 3. In this case there are ties that result from circularity in the respondents paired choices.

mdpref.out

This Is the Second Trial for MDPREF on Four Desserts

# of Subjects	# of Stimuli	# of Factors	Plotted Factors	Data Form
7	4	2	2	2

MEAN OF THE RAW SCORES (BY SUBJECT)

2.500	2.500	2.500	2.500	2.500	2.500	2.500
-------	-------	-------	-------	-------	-------	-------

SD OF THE RAW SCORES (BY SUBJECT)

0.500	1.118	1.118	1.118	0.500	1.118	1.118
-------	-------	-------	-------	-------	-------	-------

FIRST SCORE MATRIX

SUBJECT	STIMULUS			
1	-1.000	1.000	1.000	-1.000
2	-0.447	-1.342	1.342	0.447
3	-0.447	-1.342	1.342	0.447
4	-1.342	-0.447	0.447	1.342
5	1.000	-1.000	-1.000	1.000
6	1.342	-1.342	0.447	-0.447
7	1.342	-1.342	-0.447	0.447

CROSS PRODUCT MATRIX OF SUBJECTS

3	0.000	4.000	4.000	2.400	0.000	1.600	0.800
4	0.000	2.400	2.400	4.000	0.000	-1.600	-0.800
5	-4.000	0.000	0.000	0.000	4.000	1.789	3.578
6	-1.789	1.600	1.600	-1.600	1.789	4.000	3.200
7	-3.578	0.800	0.800	-0.800	3.578	3.200	4.000

CORRELATION MATRIX OF SUBJECTS

1	1.000	0.000	0.000	0.000	-1.000	-0.447	-0.894
2	0.000	1.000	1.000	0.600	0.000	0.400	0.200
3	0.000	1.000	1.000	0.600	0.000	0.400	0.200
4	0.000	0.600	0.600	1.000	0.000	-0.400	-0.200
5	-1.000	0.000	0.000	0.000	1.000	0.447	0.894
6	-0.447	0.400	0.400	-0.400	0.447	1.000	0.800
7	-0.894	0.200	0.200	-0.200	0.894	0.800	1.000

CROSS PRODUCT MATRIX OF STIMULI

1	7.800	-3.800	-3.800	-0.200
2	-3.800	9.400	-1.800	-3.800
3	-3.800	-1.800	6.200	-0.600
4	-0.200	-3.800	-0.600	4.600

ROOTS OF THE FIRST SCORE MATRIX

13.569	9.905	4.526	0.000
--------	-------	-------	-------

SECOND SCORE MATRIX

SUBJECT	STIMULUS			
1	-0.700	0.586	0.339	-0.226
2	-0.284	-0.661	0.606	0.340
3	-0.284	-0.661	0.606	0.340
4	-0.547	-0.388	0.706	0.229

(Title)

(Persons, stimuli, factors, factors plotted, standardize)

(Mean = $\sum X/N$)

(Square roots of $(\sum (X - \text{Mean})/N)$)

(Standardized (Z) scores of original preference data)

(Covariances = $\sum Z_X Z_Y$)

(Correlations = $r = \sum Z_X Z_Y / (N)$)

(Covariances, basis for determining item dimensions)

(Eigenvalues, variance accounted for by item factors)

(Captured values from preference vector fitting.
Attempts to mirror First Score Matrix. Values are
distances from the origin.)

5	0.700	-0.586	-0.339	0.226
6	0.546	-0.766	-0.104	0.324
7	0.650	-0.665	-0.252	0.268

POPULATION MATRIX

(Subject factor loadings used in plotting points)

FACTOR

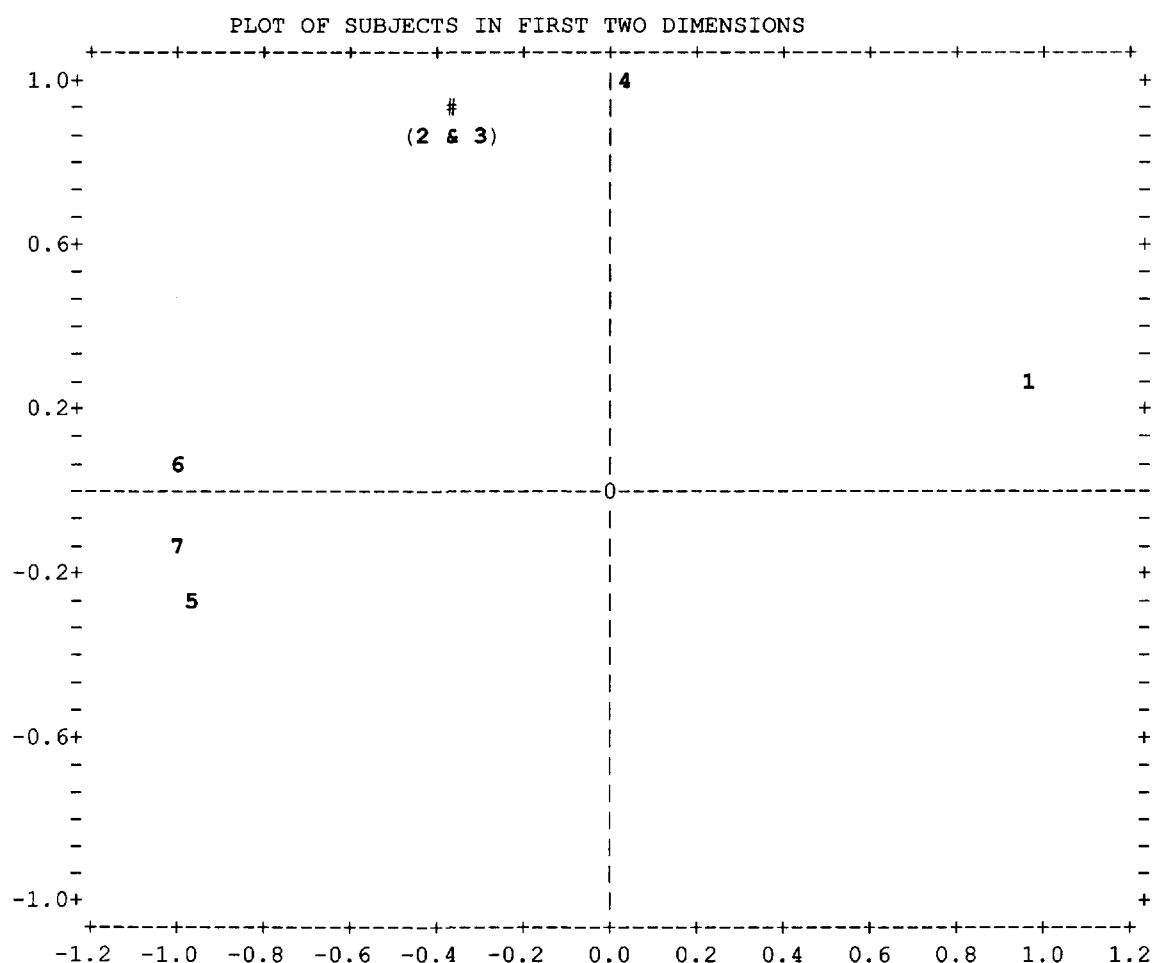
1	0.955	-0.354	-0.354	0.049	-0.955	-0.999	-0.986
2	0.297	0.935	0.935	0.999	-0.297	0.049	-0.166

STIMULUS MATRIX

(Item factor loadings. In order: Chocolate Cake, Pound Cake, Vanilla Ice Cream, Chocolate Ice-Cream)

FACTOR

1	-0.571	0.746	0.138	-0.313
2	-0.520	-0.425	0.700	0.245

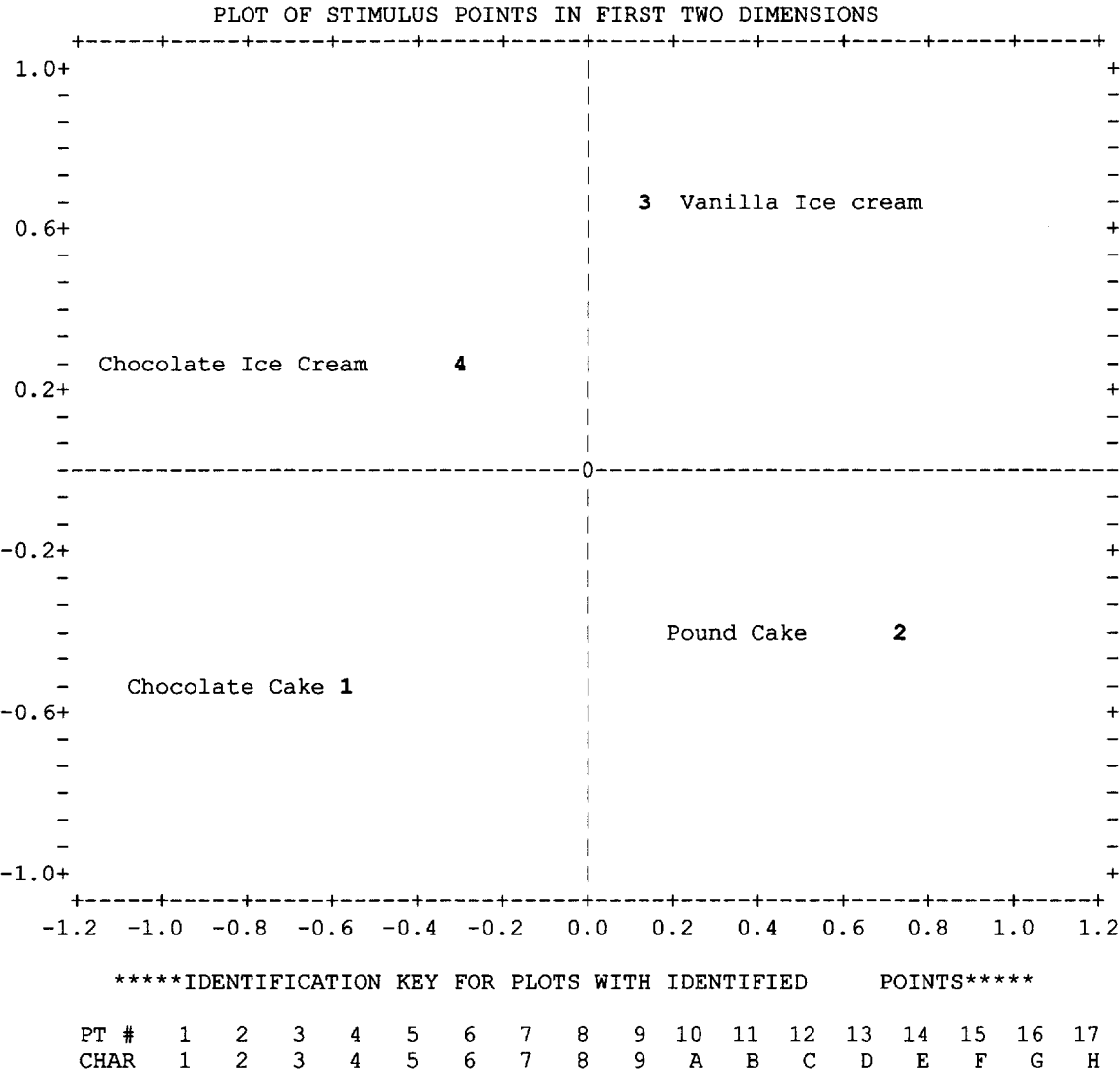


*****IDENTIFICATION KEY FOR PLOTS WITH IDENTIFIED POINTS*****

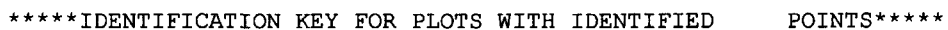
PT #	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
CHAR	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F	G	H

MULTIPLE POINTS IDENTIFIED AS # (subjects 2 and 3)

In the plot above one can see that different individuals have fairly varied positions the two dimensional space of ice cream versus cake and chocolate versus vanilla.



In this plot the dimensions of chocolate versus vanilla and cake versus ice cream are captured.



FIRST 4 POINTS ARE STIMULI AND NEXT 7 POINTS ARE SUBJECTS

In this plot both subjects and stimuli are presented. The preference vector for subject 1 (item 5 in the plot) is shown. This respondent likes vanilla as opposed to chocolate. One can see that the projections match the values in the second score matrix. These values are distances measured from the origin. For this subject, who had tied scores, the match with the original scores is only approximated.

Application : Occupational Ranking by Japanese

Hiraki (1974) studied the relative status of teachers in Japan and in Hawaii by having native Japanese visitors rank sets of occupations as to prestige. She then compared these ranks with local respondents. Hiraki used a Balanced Incomplete Block design (BIB) with 21 selected occupations ranked in groups of five (Appendix B, Table B). Figure 13.4 shows her instrument and Table 13.1 reports the results of the votes for the Japanese tourists. The Japanese vote vectors were analyzed by **MDPREF** and the two-space configuration shown as Fig. 13.5 was the result.

Hiraki also scaled the preference vectors for 21 occupations by Japanese tourists and by Americans of Japanese ancestry.

The occupations were arranged according to Plan 13.13 of Table B in the Appendix. The 21 occupations were as follows:

- | | | | |
|--------------|------------------|--------------------|-----------------|
| 1. physician | 6. lawyer | 11. union official | 16. policeman |
| 2. governor | 7. artist | 12. columnist | 17. barber |
| 3. professor | 8. factory owner | 13. electrician | 18. fisherman |
| 4. banker | 9. captain | 14. bookkeeper | 19. singer |
| 5. priest | 10. teacher | 15. farmer | 20. taxi driver |
| | | | 21. janitor |

Directions: In each section, rank order the occupations. Place (1) beside the most prestigious job (2) the next most prestigious and so on until five have been ranked.

SECTION 1	SECTION 6	SECTION 11	SECTION 16	SECTION 21
___ farmer	___ army captain	___ janitor	___ physician	___ taxi driver
___ artist	___ professor	___ lawyer	___ policeman	___ farmer
___ bookkeeper	___ governor	___ factory owner	___ nightclub singer	___ barber
___ factory owner	___ nightclub singer	___ barber	___ fisherman	___ union official
___ policeman	___ farmer	___ army captain	___ janitor	___ physician
SECTION 2	SECTION 7	SECTION 12	SECTION 17	
___ artist	___ professor	___ lawyer	___ policeman	
___ banker	___ electrician	___ columnist	___ priest	
___ army captain	___ janitor	___ physician	___ taxi driver	
___ physician	___ taxi driver	___ bookkeeper	___ governor	
___ priest	___ artist	___ professor	___ lawyer	

SECTION 3	SECTION 8	SECTION 13	SECTION 18
___banker	___electrician	___columnist	___priest
___barber	___fisherman	___union official	___teacher
___professor	___lawyer	___policeman	___farmer
___policeman	___farmer	___army captain	___janitor
___teacher	___banker	___fisherman	___columnist
SECTION 4	SECTION 9	SECTION 14	SECTION 19
___barber	___fisherman	___union official	___teacher
___bookkeeper	___governor	___physician	___nightclub singer
___electrician	___columnist	___nightclub singer	___artist
___priest	___artist	___fisherman	___lawyer
___nightclub singer	___barber	___janitor	___union official
SECTION 5	SECTION 10	SECTION 15	SECTION 20
___bookkeeper	___governor	___phsysician	___nightclub singer
___army captain	___janitor	___policeman	___taxi driver
___fisherman	___union official	___nightclub singer	___banker
___teacher	___banker	___fisherman	___columnist
___taxi driver	___bookkeeper	___janitor	___factory owner

FIG 13.4. Example of a BIB instrument used to evaluate occupations.

Table 13.1

Occupational Votes (Japan)

1	20	17	15	16	9	17	11	17	9	8	10	11	14	8	6	6	3	3	2	4	4
2	16	2	13	11	5	16	20	16	3	9	2	19	5	0	18	10	11	11	8	9	6
3	15	9	8	9	14	19	11	6	0	15	2	12	18	5	20	8	15	12	1	6	5
4	12	14	14	12	6	9	7	14	7	9	6	7	14	15	14	7	13	7	6	10	7
5	20	16	15	14	3	18	11	19	12	6	9	10	13	3	17	3	8	0	4	6	3
6	13	7	11	2	0	14	16	6	4	8	14	12	17	1	19	3	9	20	18	11	5
7	19	13	17	13	13	17	11	15	6	3	7	20	11	3	16	7	6	8	1	1	3
8	15	11	5	6	15	19	1	12	9	11	5	15	13	12	14	6	9	5	14	11	2
9	16	20	17	7	18	19	13	9	0	13	6	14	11	11	13	6	4	7	1	3	2
10	15	20	11	10	8	17	10	19	18	6	14	12	12	5	16	1	7	4	0	3	2
11	4	17	17	3	13	7	12	16	17	1	0	20	18	7	12	6	11	9	11	3	6
12	19	11	19	16	14	18	12	12	12	14	6	9	17	8	6	6	4	3	0	3	1
13	20	19	16	10	15	18	7	14	12	5	17	11	13	3	9	8	4	5	3	1	0
14	17	10	5	3	18	15	14	1	0	11	11	13	10	3	20	7	10	19	7	5	11
15	15	17	12	10	15	18	11	19	5	11	7	11	18	6	17	5	5	4	3	1	0

16	17	17	17	13	7	18	14	13	7	15	18	5	9	10	8	8	5	2	6	1	0
17	19	14	18	16	8	16	8	20	10	10	12	12	13	8	4	4	2	0	2	4	
18	15	11	6	9	10	13	2	9	3	3	5	5	17	7	20	11	14	18	1	14	
19	16	19	20	13	12	18	10	14	4	17	6	5	8	6	14	12	8	2	0	5	
20	20	13	11	8	16	19	8	18	0	16	4	9	15	7	13	8	5	10	3	4	
21	18	15	12	18	5	17	10	17	10	1	10	11	11	6	20	3	7	9	8	1	
22	14	7	12	11	6	10	7	13	1	15	20	16	18	6	19	0	9	2	3	4	
23	9	20	14	8	7	15	14	15	19	14	6	9	13	3	10	9	1	1	19	4	
24	16	19	15	8	10	18	14	17	12	13	5	9	11	5	20	6	5	2	0	1	
25	19	20	14	13	4	18	6	17	12	9	8	15	12	10	7	11	7	3	2	1	
26	20	16	11	17	4	13	4	17	16	10	9	3	11	12	19	9	8	2	0	6	
27	17	19	15	6	6	20	1	11	13	15	7	10	18	9	15	12	7	3	0	4	
28	17	20	15	15	12	18	9	19	0	11	12	7	8	5	12	13	5	5	1	4	
29	18	16	19	6	14	20	15	12	16	10	10	6	10	3	11	11	2	6	4	1	
30	20	11	19	10	10	17	4	17	5	14	4	10	14	17	13	5	7	1	1	9	
31	15	20	17	13	9	18	12	19	4	16	13	8	11	5	11	4	6	2	0	3	
32	15	20	16	13	6	19	17	17	9	7	11	11	6	10	7	4	4	0	15	1	
33	17	13	12	14	12	15	15	20	2	8	12	19	8	3	13	1	6	10	1	4	
34	17	17	18	13	20	17	16	13	5	13	9	9	5	4	10	9	3	5	0	3	
35	20	13	12	18	0	19	5	15	17	9	8	11	14	14	12	5	2	5	7	3	
36	18	9	12	13	7	16	19	13	0	9	8	11	16	5	19	6	13	8	1	3	
37	19	9	20	14	11	18	12	10	12	15	5	10	17	8	10	2	4	3	4	6	
38	17	10	12	19	16	14	15	20	2	12	0	8	8	6	18	5	8	11	1	3	
39	17	19	13	8	5	18	8	20	16	12	6	9	15	13	12	1	9	1	4	1	
40	14	17	15	15	9	19	19	17	6	6	2	5	11	5	11	2	10	8	12	4	
41	19	15	13	18	9	18	12	19	10	9	3	12	6	5	14	6	4	2	14	1	
42	15	17	15	4	17	18	12	19	6	14	1	12	5	3	10	5	6	11	10	5	
43	16	7	19	4	20	14	10	8	11	18	6	5	11	3	12	6	11	14	10	3	
44	19	13	16	7	9	16	10	11	6	20	6	3	16	5	14	18	11	0	1	6	
45	9	16	18	3	1	17	7	16	1	13	17	20	15	8	12	7	10	5	4	7	
46	16	18	12	7	6	17	3	19	1	9	5	16	14	12	19	6	10	11	0	6	
47	20	7	9	9	7	11	2	6	0	11	10	8	18	17	14	6	15	19	1	6	
48	13	18	18	14	7	18	16	16	11	10	5	9	12	5	17	9	6	3	0	1	
49	2	3	4	8	20	5	18	1	0	7	10	11	12	13	6	9	17	19	14	15	
50	15	14	9	19	9	17	11	20	7	5	4	14	7	14	18	3	4	12	3	0	

Output MDPREF Preferred Occupations by Japanese Tourists

For this analysis only the object and subject factor coordinates are provided below.

POPULATION MATRIX

FACTOR												
1	0.974	0.709	0.546	0.995	0.988	0.146	0.964	0.960	0.941	0.968		
	0.916	0.992	0.982	0.172	0.985	0.943	0.968	0.073	1.000	0.893		
	1.000	0.751	0.665	1.000	0.947	0.956	0.998	0.998	0.972	0.997		
	0.999	0.848	0.941	0.997	0.904	0.819	0.998	0.883	0.951	1.000		
	0.963	0.983	0.841	0.986	0.987	0.897	0.174	0.995	-0.809	0.987		
2	0.226	-0.705	-0.838	-0.103	0.153	-0.989	-0.264	-0.278	-0.338	0.250		
	0.400	0.130	0.186	-0.985	-0.170	0.333	0.249	-0.997	-0.027	-0.450		
	0.001	-0.660	0.747	0.007	0.321	0.294	0.068	-0.057	0.237	-0.075		
	0.054	0.530	-0.339	-0.072	0.428	-0.574	0.071	-0.469	0.310	0.025		
	0.270	-0.182	-0.541	-0.165	-0.158	-0.442	-0.985	0.098	-0.588	-0.161		

STIMULUS MATRIX

[illegible]

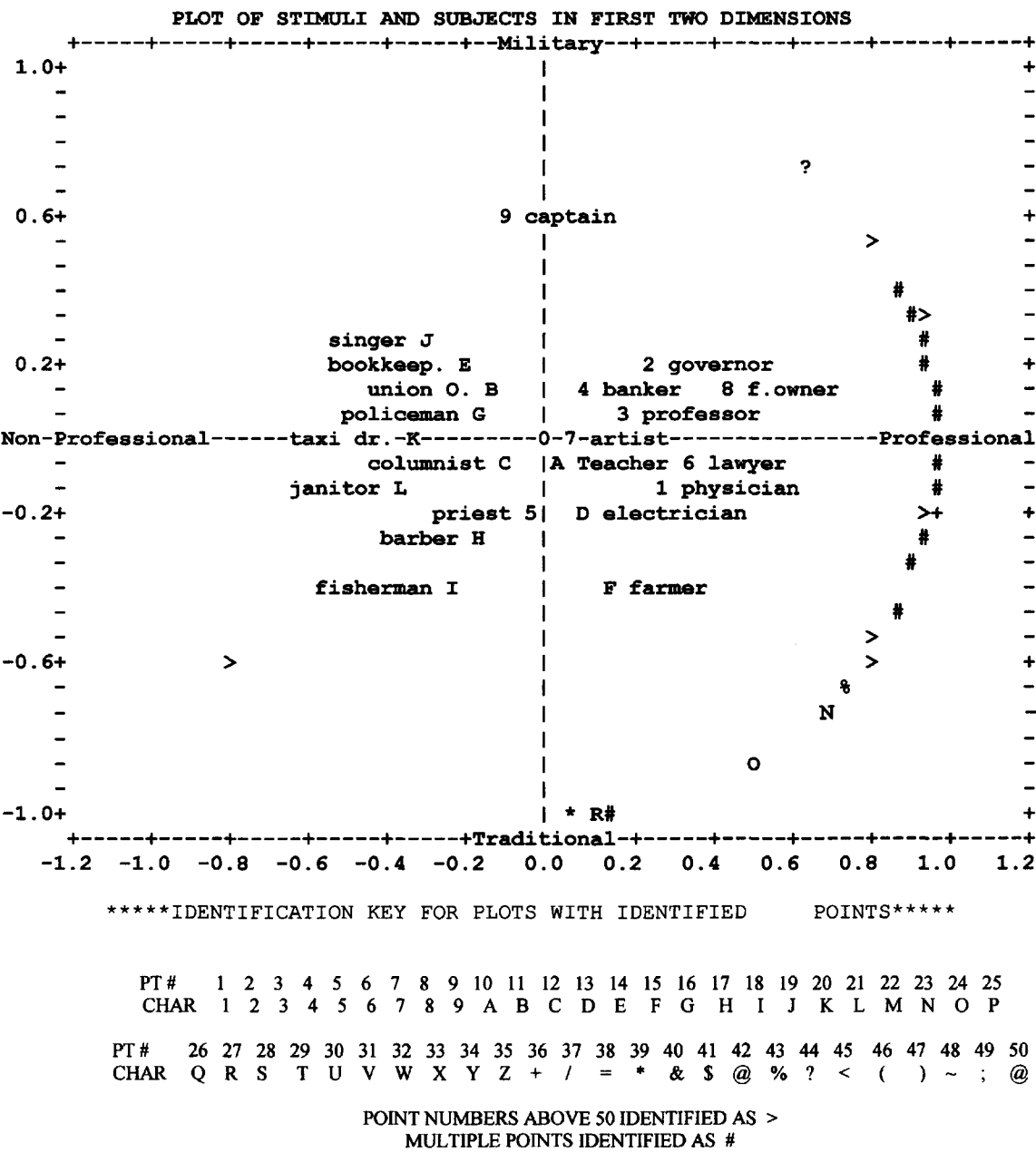


FIG. 13.5. Two dimensional configuration of occupational votes by Japanese tourists using MDPREF.

Hiraki was able to show that native Japanese visitors neatly divide occupations into professional and less prestigious categories. Only a few respondents indicated a preference for the traditional occupation of fisherman but farmer is still desired. These occupations are contrasted with the military occupation of captain.

Inclusion of the Ideal Point

A simple way to measure individual preference is to include an "ideal" stimulus among the authentic stimuli and obtain similarity estimates among all the $(n + 1)$ stimuli. If, for example, the ideal professor is included among the names of the graduate faculty and similarity estimates among faculty members are obtained from each graduate student in a department, it is assumed that those professors scaled closest to the student's ideal professor are most preferred. The scaling is done using multidimensional methods for each student.

Ideal Point Projection

If we can find the most preferred combination of attributes on the map of the stimuli, this would be the ideal point. The distances to the ideal point from the stimuli should closely match the preference scale values. The differences between these distances and the scale values are minimized and the location of the ideal point is determined. Finding a respondent's ideal point is analogous to finding the vector solution. Suppose a child's preference ranking for the four desserts chocolate cake, chocolate ice cream, pound cake and vanilla ice cream is 1, 2, 3, 4. Using the ranks as distances we can project them onto an ideal point as follows:

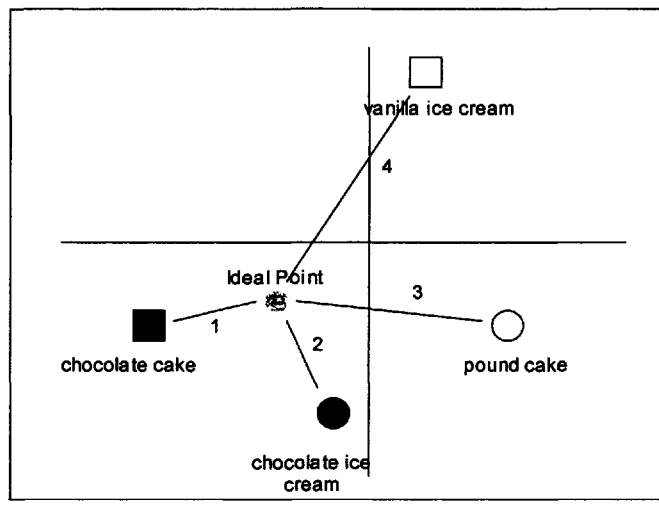


FIG. 13.6. Ideal point projection.

By placing Child A's ideal point in the position shown in Fig 13.6, the distances closely approximate the preference values. Carroll and Chang (1968) produced the program PREFMAP which attempts to locate the ideal points for subjects in the space of the Stimuli. This program is available from BYU or Bell Labs netlib.

14

MULTIDIMENSIONAL SCALING

Multidimensional Scaling (MDS) is the name for specific methods that attempt to spatially represent the proximities between a number of stimuli. MDS attempts to provide a picture of the similarities between objects by mapping the distances between them. The MDS methods are alternatives to cluster and factor analysis. Each object is represented as a point in space (usually in the Euclidean space of two or three dimensions but could include hyperspace). The distances between the objects inversely represent the similarities. Two objects close together are highly similar. Two objects far apart are dissimilar. A brief historical introduction to MDS is given by Davison (1983) and Young (1985)

Multidimensional Scaling is applicable to a large number of measures of similarity or dissimilarity. Unlike factor analysis, MDS can be used on data with fewer assumptions about the data. Its primary purpose is a parsimonious spatial representation of the objects. Shepard (1962) first showed that MDS could utilize ordinal (nonmetric) assumptions about the data and produce metric solutions.

The nonmetric process of arriving at the best spatial configuration to represent the original similarities has been presented by Kruskal (1964a). This multidimensional scaling proceeds as follows:

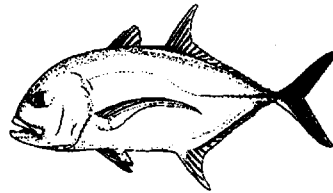
1. There is a given set of (n) objects.
2. For every two objects (i and j), some measure or function of proximity is obtained. (These measures may be correlations, similarities, associations, etc.). Distance and similarity measures are related inversely. If, for example, the similarity between the words "war" and "peace" is estimated to be small, then the two words should be a relatively "large" distance apart. If similarities are obtained (s_{ij}), they are converted to distances (d_{ij}) usually by subtracting from a constant. An additive constant may also be needed to fulfill the requirements of the triangle inequality.
3. A number of dimensions, t , are selected that may be needed to fit the data. The n objects are then placed (randomly or selectively) in the space of the chosen number of dimensions spaces.
4. Multidimensional scaling (MDS) searches for a replotting of the n objects so that (created) distances (d^*_{ij}) between pairs of objects in the plot are related to their original distances (d_{ij}).

In Kruskal's method, a resolution of the initial spatial configuration is made in iterative steps. At each step the objects are moved from their initial placement in the t dimensional space (usually 2 or 3 dimensions) and new distances between all pairs of objects are calculated. The distances (d_{ij}^*) between pairs of objects in the new placement are ordered and then compared with the original dissimilarities (d_{ij}) between the same pairs of objects. The original distances have also been ordered. If the relationship between the two sets of ranks is increasingly monotonic, that is, if the order of the new distances is similar to the order of the original distances, the objects continue to move in the same direction at the next step. If the relationship is not monotonic, changes in direction and step length are made. A measure of monotonicity is primary in non-metric scaling. This measure is provided by ordering the distance measures (d_{ij}^*) on the x-axis and measuring horizontal deviations of the newly obtained distances in the plot from the original distances (d_{ij}) on the y-axis. The deviations are squared so they can be summed. The object is to make the sum of the squared deviations as small as possible. That is, to make $\sum (d_{ij} - d_{ij}^*)^2$ a minimum, which is a typical least squares problem.

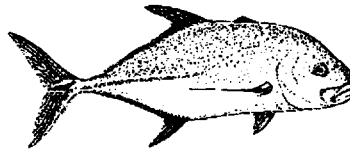
How Kruskal's Method Works

Suppose the similarities (s_{ij}) among four fish (all pacific trevallys) are obtained as an average of several fisherman's subjective estimates.

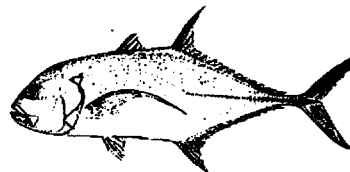
Giant Trevally



Bluefin Trevally



Brassy Trevally



Bigeye Trevally



The fisherman’s average similarities are formed into a square matrix, Table 14.1:

Table 14.1
Similarities (s_{ij}) Among Trevallys

	Giant	Bluefin	Brassy	Bigeye
Giant	10	6	3	5
Bluefin	6	10	7	4
Brassy	3	7	10	4
Bigeye	5	4	4	10

These are converted to distances (d_{ij}) by subtracting each similarity score from 10.

Table 14.2
Distances (d_{ij}) Between Trevallys

	Giant	Bluefin	Brassy	Bigeye
Giant	0			
Bluefin	4	0		
Brassy	7	3	0	
Bigeye	5	6	6	0

No additive constant needed.

As an initial step, the four fish are placed, **randomly** as points in a two-dimensional plane ($t = 2$). Euclidean distances, d^*_{ij} between all pairs, are determined from the initial placement coordinates.

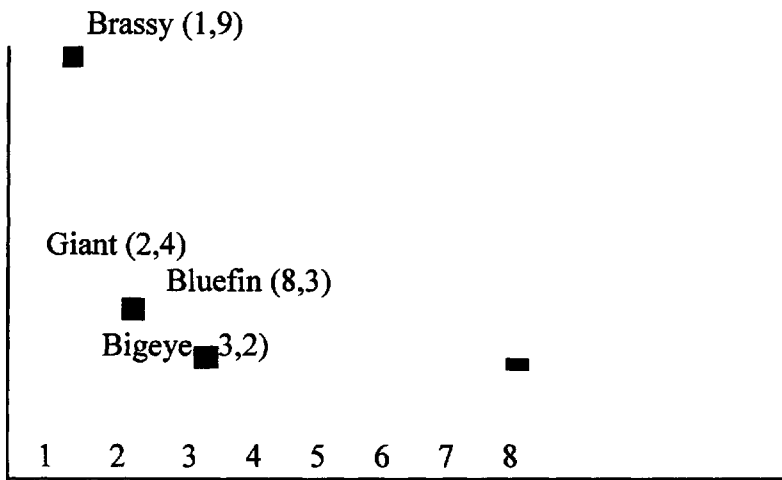


FIG. 14.1. Random placement of fish in two dimensions.

Because the rank of the original distances and the rank of the plotted distances are not in the same order, the first spatial representation based on random assignment does not fit the assumption of monotonicity (that is, steadily increasing or decreasing values).

Table 14.3
Distances Between Pairs of Reef Predators

	d_{ij}	d_{ij}^2	Rank	d^*_{ij}	Rank	$D = (d_{ij} - d^*_{ij})$	D^2
Bluefin-Brassy	3	9	1	5.1	2.5	-2.1	4.41
Bluefin-Giant	4	16	2	6.1	4	-2.1	4.41
Giant-Bigeye	5	25	3	2.2	1	2.8	7.84
Bluefin-Bigeye	6	36	4.5	5.1	2.5	0.9	.81
Brassy-Bigeye	6	36	4.5	7.3	5	-1.3	1.69
Giant-Brassy	7	49	6	9.2	6	-2.2	4.84
	152			Sum	$D^2 = \text{RawStress}$		13.0

As one measure of how well the random point placement fits the original data, a calculation of Raw Stress, which is the sum $D^2 = \sum (d_{ij} - d^*_{ij})^2$ can be made. In this case, d_{ij} are the original distances and d^*_{ij} are the distance estimates based on random assignment. Kruskal averages the Raw Stress Sum of Squares by dividing by the Sum of d_{ij}^2 and then getting back to the original units by taking the square root. In this example $13 / 152 = .086$ and the square root is .29.

One can also find the correlation (r) between the ranks. One measure of fit is the square root of $(1 - r^2)$ sometimes called the **badness of fit**. The smaller the value, the better the fit to the original data. r^2 is the proportion of variance accounted for and $(1 - r^2)$ is the coefficient of nondetermination.

As a next step, the four points are moved varying small amounts in distance and direction so as to reduce the stress index or badness of fit index. In the example above, the Giant Trevally and the Bigeye Trevally should be moved away from each other and the Bluefin-Brassy pair and the Bluefin-Giant pair should be moved closer together.

In practice, Kruskal's method is one of successive approximation. Theoretically the problem is to minimize a stress function involving many variables in t dimensions. The solution of the problem is found in the method of steepest descent. See Kruskal (1964b) and Kruskal and Wish (1978). In current versions of MDS, principal components analysis is used to determine the **initial** placement of the points in space.

Using SAS MDS the original similarities were analyzed: SAS detects whether the data are similarities or distances by comparing the diagonal elements with largest off diagonal values. The reader may also download (kyst2a.f) from the Bell-Labs net library as well as the manual. (See Using the Internet, p. 208).

SAS Analysis of Trevally Data

```
data trevally;
input Bluefin Giant Bigeye Brassy;
datalines;
10 6 4 7
6 10 5 3
4 5 10 4
7 3 4 10
;
proc mds data=trevally dim=2 out=out level=ord
pfinal;
proc plot data=out;
plot dim2*dim1;
run;
```

}Input to SAS

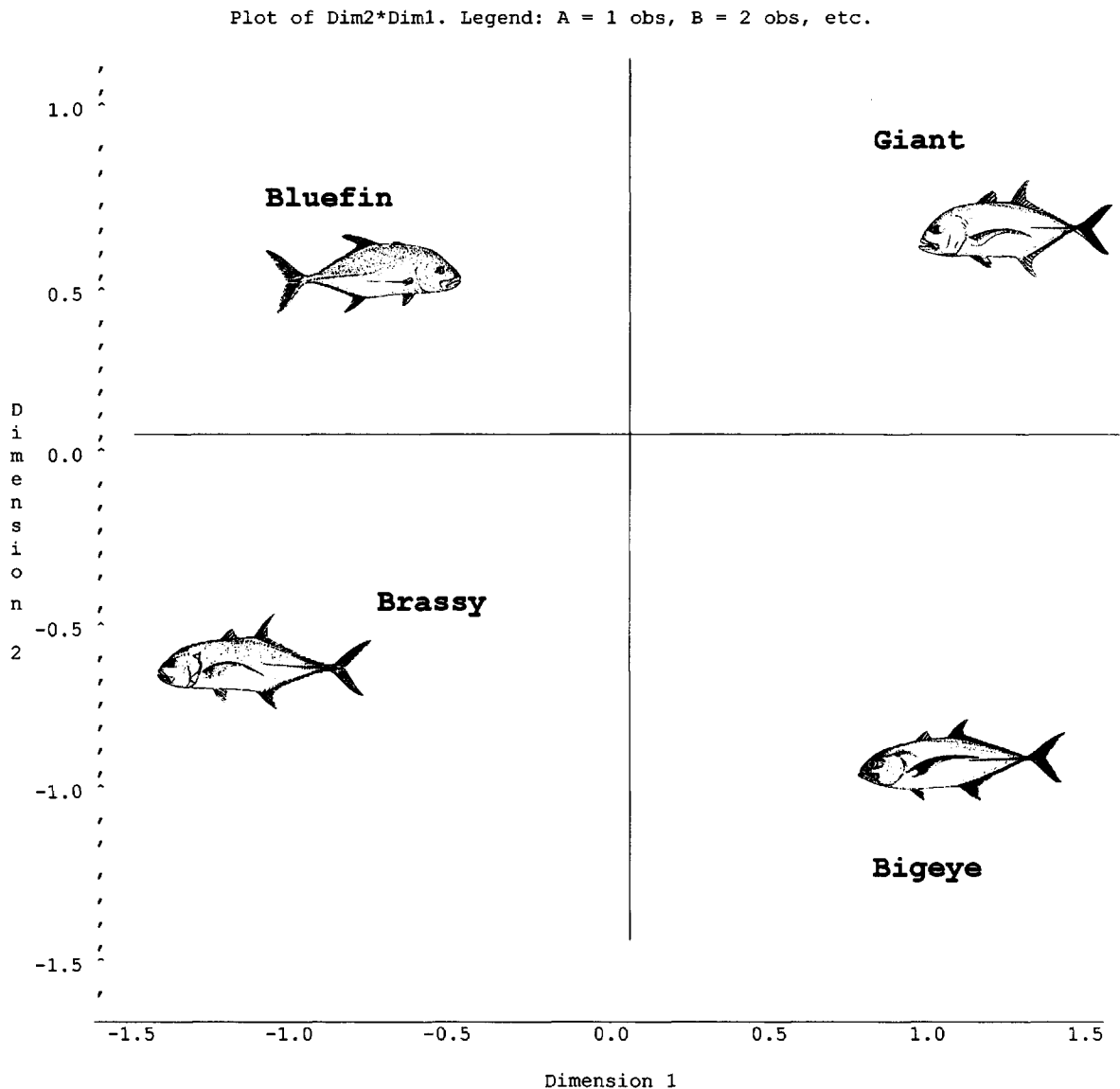
Multidimensional Scaling: Data=WORK.TREVALLY
Shape=TRIANGLE Condition=MATRIX Level=ORDINAL
(Shape = TRIANGLE means the upper triangle of the symmetric matrix was used.)
Coef=IDENTITY Dimension=2 Formula=1 Fit=1 (data are distances)
(Coef = IDENTITY means principal axis rotation and scaled to a standard deviation of 1)
(Formula = 1 means Kruskal's Stress formula 1)

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

Iteration	Type	Badness- of-Fit Criterion	Change in Criterion	Convergence Measures	
				Monotone	Gradient
0	Initial	0.0391	.	.	.
1	Monotone	0.002918	0.0362	0.0325	0.8852
2	Gau-New	0.001365	0.001552	.	.
3	Monotone	0.000632	0.000733	0.001210	0.8825
4	Gau-New	0.000297	0.000334	.	0.003527

Convergence criteria are satisfied.

Configuration		
	Dim1	Dim2
Bluefin	-0.65	0.76
Giant	1.17	0.90
Bigeye	0.95	-1.25
Brassy	-1.48	-0.42



If one physically measures the distances between the trevallids in the plot, the final distances (d_{ij}) match the original distances very closely. For fisherman, the arrangement makes sense. The bluefin and brassy trevallids are shallow water reef predators whereas the giant and bigeye trevallids are found in deeper water. The brassy and bigeye trevallids are more common south of the equator whereas the giant and bluefin's habitat is wide ranging.

Application Word Similarity (SAS MDS Using PEROVER Data)

Fifteen adults were given 11 words on slips of paper. Each word began with the letter **a**. The subjects were asked to group similar words together. A percent overlap matrix was calculated and a distance matrix produced. (See Free Clustering, p. 23).

SAS Data Set

```

data words;
input a admits aged almost aiming and as at areas army away;
datalines;
0.00 1.00 0.93 1.00 1.00 0.73 0.27 0.33 1.00 1.00 1.00
1.00 0.00 0.93 0.40 0.67 0.93 1.00 1.00 0.87 0.80 0.80
0.93 0.93 0.00 0.93 0.93 0.80 0.93 0.93 0.67 0.87 0.87
1.00 0.40 0.93 0.00 0.60 1.00 1.00 1.00 0.93 0.80 0.80
1.00 0.67 0.93 0.60 0.00 1.00 1.00 1.00 0.87 0.73 0.73
0.73 0.93 0.80 1.00 1.00 0.00 0.87 0.87 0.73 0.87 0.80
0.27 1.00 0.93 1.00 1.00 0.87 0.00 0.27 1.00 1.00 1.00
0.33 1.00 0.93 1.00 1.00 0.87 0.27 0.00 1.00 1.00 1.00
1.00 0.87 0.67 0.93 0.87 0.73 1.00 1.00 0.00 0.73 0.73
1.00 0.80 0.87 0.80 0.73 0.87 1.00 1.00 0.73 0.00 0.13
1.00 0.80 0.87 0.80 0.73 0.80 1.00 1.00 0.73 0.13 0.00
;
proc print;
proc mds dim=2 out=out pfinal;
proc plot data=out;
plot dim2*dim1

```

} SAS Data Step

} Data = Distance Matrix

} Analysis

The results are listed below:

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

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

		Convergence Measures			
	Iteration Type	Badness- of-Fit Criterion	Change in Criterion	Monotone	Gradient
0	Initial	0.2560			
1	Monotone	0.1372	0.1188	0.2176	0.7605
2	Gau-New	0.0842	0.0530		
3	Monotone	0.0809	0.003358	0.0229	0.4051
4	Gau-New	0.0793	0.001596		
5	Monotone	0.0723	0.006932	0.0328	0.3139
6	Gau-New	0.0711	0.001251		
7	Monotone	0.0699	0.001149	0.0135	0.2474
8	Gau-New	0.0693	0.000663		
9	Monotone	0.0688	0.000457	0.007731	0.2031
10	Gau-New	0.0675	0.001362		0.0197
11	Gau-New	0.0674	0.0000120		0.002906

Convergence criteria are satisfied.

	Configuration	
	Dim1	Dim2
a	1.59	0.35
admits	-1.01	0.94
aged	-0.09	-1.65
almost	-1.01	1.12
aiming	-1.14	0.74
and	0.76	-0.86
as	1.61	0.53
at	1.58	0.61
areas	-0.41	-1.20
army	-0.99	-0.29
away	-0.89	-0.30

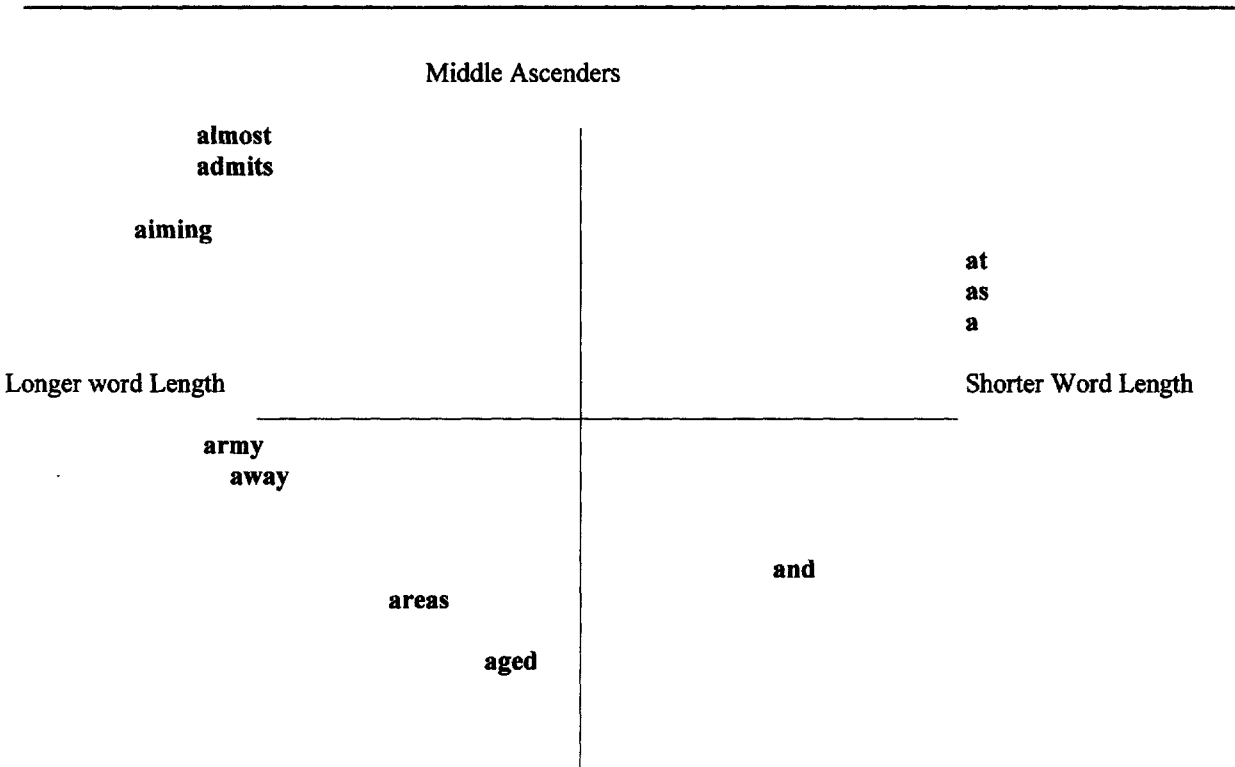


FIG. 14.2. Similarity between words beginning with the letter a.

In this example, one could argue that the horizontal dimension is one of word length. The vertical dimension is characterized by words with middle ascenders versus words without. The badness of fit criterion ends at .0675 which is good and convergence was reached. Note this solution was nonmetric in that only ordinal conditions were imposed on the solution.

The cluster analysis of this same data can be super-imposed on the MDS solution as shown below:

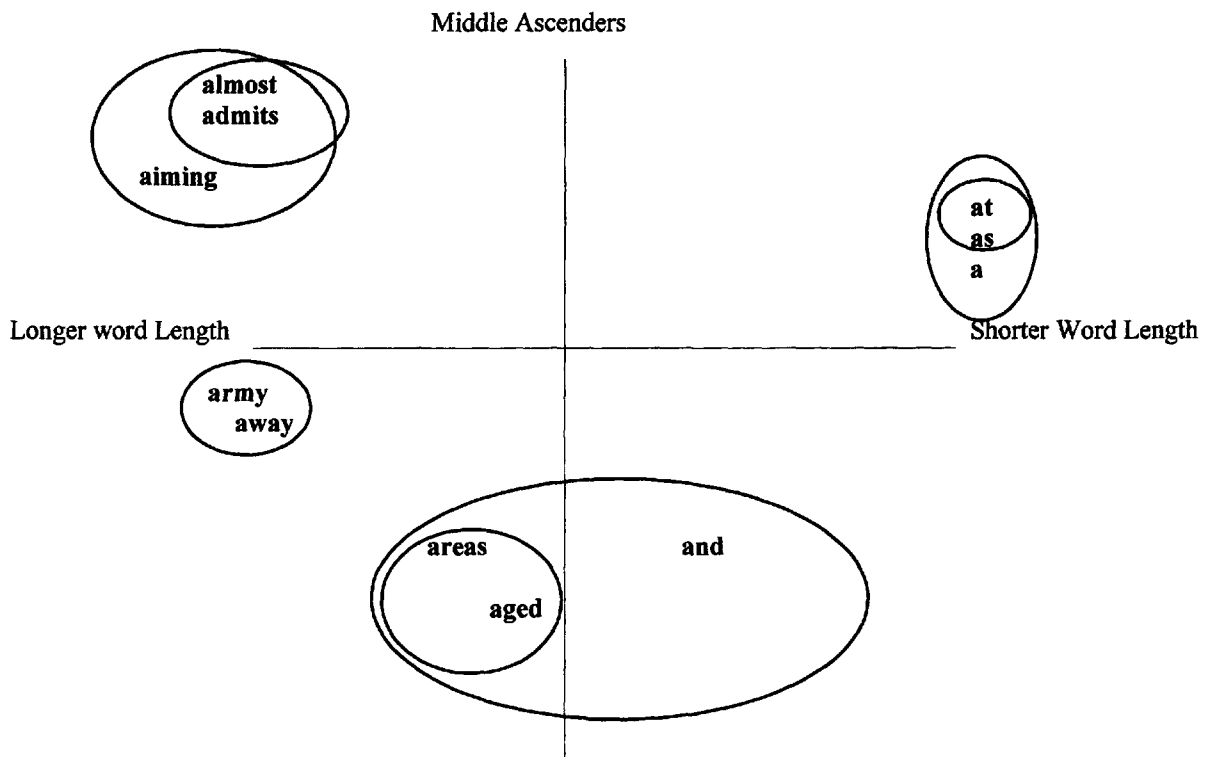


FIG. 14.3. Superimposing clustering on the MDS analysis.

The clustering emphasizes word length, creating clusters of small, medium, and large words. The word **and** stretches its cluster because some judges see it as a small word.

This page intentionally left blank

15

INDIVIDUAL DIFFERENCES SCALING

Individual Differences Scaling is a form of weighted multidimensional scaling. Its basic assumption is that each individual responds to all the dimensions of the stimuli but may utilize the dimensions in varying degrees. That is, each subject weights the dimensions separately and perhaps differently. The classic example is people who may be color blind in the red-green dimension of the color wheel. For such subjects, the red-green dimension has a minimal weight. It can be seen that a separate matrix of information is required from each subject. The auxillary program INDMAT accepts paired similarity data and produces a set of stacked matrices. For example:

Output from INDMAT

7.00	5.00	2.00	4.00
5.00	7.00	2.00	2.00
2.00	2.00	7.00	5.00
4.00	2.00	5.00	7.00
7.00	6.00	2.00	4.00
6.00	7.00	3.00	2.00
2.00	3.00	7.00	7.00
4.00	2.00	7.00	7.00
.			
.			
.			
7.00	3.00	1.00	6.00
3.00	7.00	6.00	3.00
1.00	6.00	7.00	4.00
6.00	3.00	4.00	7.00

or

Judge 1			
7	5	2	4
5	7	2	2
2	2	7	5
4	2	5	7

Judge 2			
7	6	2	4
6	7	3	2
3	3	7	7
4	2	7	7

Judge 8			
7	3	1	6
3	7	6	3
1	6	7	4
6	3	4	7

SINDSCAL

The initial individual differences methodology was developed by Carroll and Chang (1968) and promoted through the effective FORTRAN program sindscal.f. This program can be downloaded free from the Bell Telephone Laboratories Library of MDS programs (see Using the Internet). It can also be found on the **CD-ROM (SINDSCAL)**.

CD-ROM Example of SINDSCAL With Learning Disability Data

The data in which five subjects gave estimates of the similarity between Learning Disability (LD), Mentally Retarded (MR), Deaf (D), and Blind (B) is examined under SINDSCAL (see p. 5). SINDSCAL can handle a variety of input matrices. In this example, the data is a lower half matrix for each judge without the diagonal.

The first line is the title line.

The second line contains the parameters:

The minimum dimensions.	2
The maximum dimensions.	2
The number of subjects.	5
The number of objects.	4
Max number of iterations	20
Uses a lower half matrix	1
Plot the data.	0
A random starting number	3333

The third line is the input file name

The fourth line is the output file name

sindscal.cfg (Configuration file)

```
Sindscal on Disability Data
2 2 5 4 20 1 0 3333
sindscal.dat
sindscal.out
```

sindscal.dat (Input file)

```
4.00
4.00 5.00
4.00 2.00 5.00
6.00
3.00 8.00
2.00 2.00 4.00
5.00
4.00 6.00
4.00 3.00 4.00
2.00
2.00 4.00
6.00 2.00 5.00
2.00
6.00 7.00
6.00 4.00 5.00
```

sindscal.out (Output file)**SINDSCAL ON DISABILITY DATA**

PARAMETERS

DIM IRDATA ITMAX IPLOT IRN

2 1 100 0 3333

NO. OF MATRICES = 5 NO. OF STIM. = 4

INITIAL STIMULUS MATRIX

1 -0.009 -0.191 -0.496 0.349 (Randomly generated coordinates)

2 -0.100 0.366 0.490 -0.078

ITERATION	HISTORY OF COMPUTATION		LOSS (Y-YHAT)**2
	CORRELATIONS BETWEEN Y (DATA) AND YHAT	VAF (R**2)	
0	0.583882	0.340918	0.659082
1	0.854795	0.730674	0.274808
2	0.913084	0.833722	0.166281
3	0.934468	0.873230	0.126772
4	0.941055	0.885585	0.114416
5	0.941746	0.886885	0.113115
6	0.942028	0.887416	0.112584
7	0.942143	0.887633	0.112367
8	0.942154	0.887655	0.112345
9	0.942160	0.887666	0.112334
10	0.942162	0.887668	0.112331
11	0.942162	0.887669	0.112331

REACHED CRITERION ON ITERATION 11

FINAL	0.942162	0.887669	0.112331
-------	----------	----------	----------

NORMALIZED SOLUTION**SUBJECTS WEIGHT MATRIX**

1 0.633 0.454 0.606 0.872 0.882

2 0.719 0.815 0.700 0.303 0.087

STIMULUS MATRIX

1 0.576 -0.600 -0.380 0.404

2 -0.472 -0.455 0.197 0.729

NORMALIZED SUM OF PRODUCTS (SUBJECTS)

1 1.000

2 0.754 1.000

SUM OF PRODUCTS (STIMULI)

1 1.000

2 0.221 1.000

APPROXIMATE PROPORTION OF TOTAL VARIANCE ACCOUNTED FOR BY EACH DIMENSION

1 2

0.521 0.367

CORRELATION BETWEEN COMPUTED SCORES AND SCALAR PROD. FOR SUBJECTS

1 0.981081

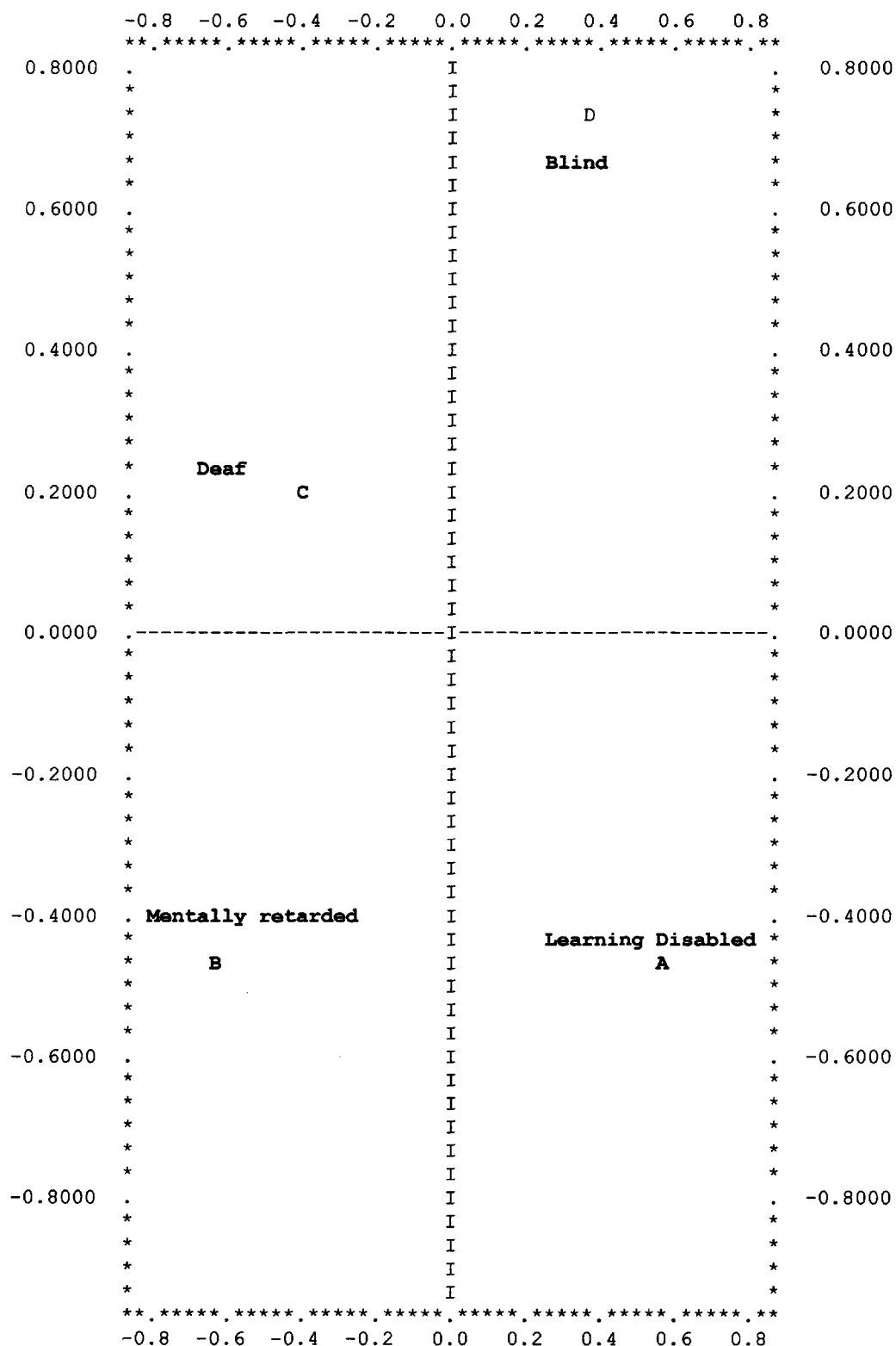
2 0.951668

3 0.947979

4 0.937011

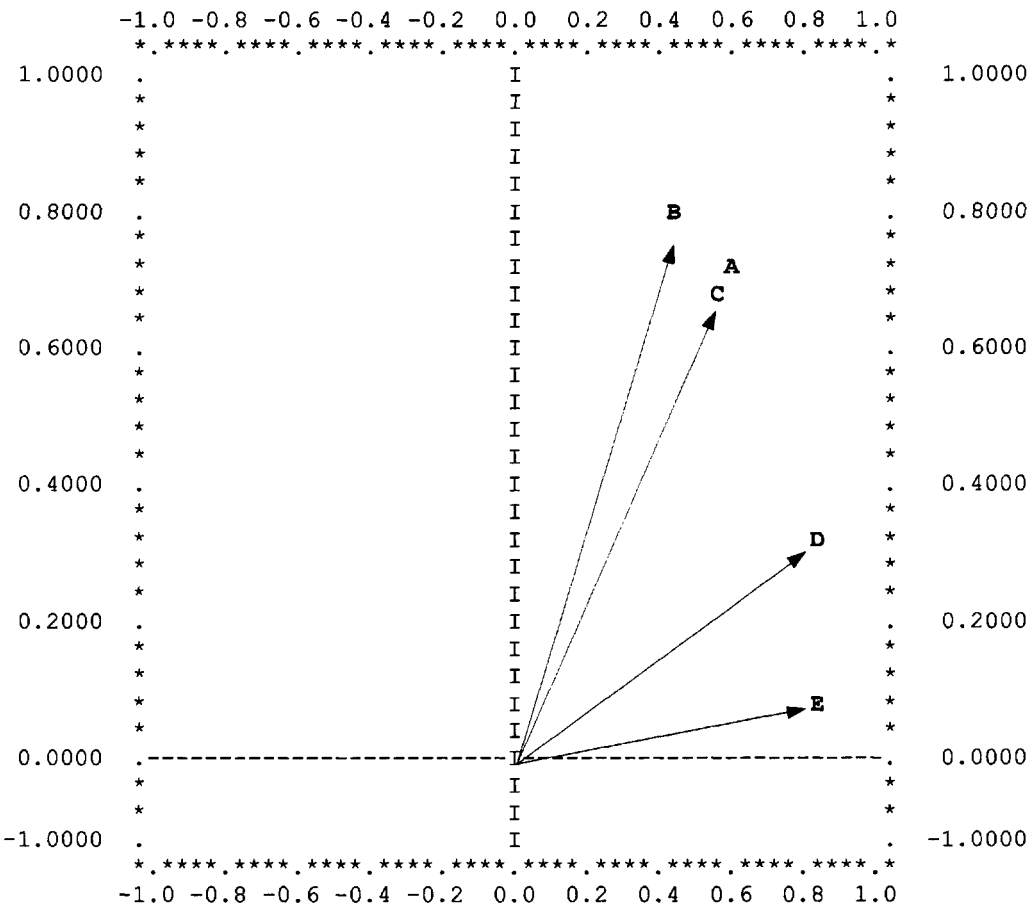
5 0.890787

STIMULUS SPACE DIMENSION 2 (Y-AXIS) VS. DIMENSION 1 (X-AXIS)



The dimensional structure indicates that the vertical or y axis can be thought of as contrasting physical versus mental disabilities. It is suggested that the horizontal or x axis is the contrast between disabilities that are capable of being mainstreamed, blind and learning disabled as opposed to more severe educational handicaps.

WEIGHTS SPACE DIMENSION 2 (Y-AXIS) VS. DIMENSION 1 (X-AXIS)



The plot indicates that subject E is making judgments of similarity primarily on the basis of mainstreaming whereas subject B is primarily using the Physical — Mental Disabilities dimension in making judgments of similarity (see plot of stimuli). The length of the lines from the origin to the plotted points reflect the amount of variance accounted for by the judgments for each subject. Note that all subjects variance is generally accountable. Subject E has the lowest correlation between computed scores and original scalar products. The scale is slightly distorted so that it will fit on regular 8 1/2 by 11 inch paper.

How SINDSCAL Works

Weighted MDS provides a group stimulus space and a weight space. The weights describe the importance each judge gives to each dimension. The closer the weight is to one (1.0) the more that dimension is used in making similarity decisions.

More formally a SINDSCAL analysis is performed on the scalar product (see p. 45) distances between objects. There is an initial configuration and created configuration. The function to be minimized is

$$\sum_i |S_i - XW_iY'|^2$$

where S_i are the known symmetric matrices of scalar products,

W_i are unknown diagonal matrices with weights (the weights are positive) and

X and Y are unknown configuration matrices which converge.

Carroll and Chang (1970) obtain convergence with an algorithm called CANDECOMP for Canonical Decomposition similar to singular value decomposition (SVD). A canonical form of a matrix is the most convenient and usually simplest form to which a square matrix can be reduced by a certain type of transformation. The canonical form has nonzero elements in the diagonal and zeros elsewhere as was demonstrated in Singular Value Decomposition (SVD).

After the scalar products are determined as well as the additive constant, if necessary, the sums of the squares between all the stimuli are determined and the objects randomly placed in (t) dimensions. The original distances d and the created \hat{d} distances are iteratively compared and adjusted, using a number of iterations or a specific R^2 as a criterion. The data are normalized and plotted.

ALSCAL

Young (Young & Lewyckyj, 1979) developed ALSCAL which stands for Alternating Least Squares Analysis which is used by SAS for classical MDS as well as weighted MDS usually known as Individual Differences Scaling, and a number of other methods. The Market Research Application (SAS, 1996) is the most user-friendly option for using ALSCAL but restricts the data to a certain form, that is, a square symmetric matrix or a set of square matrices (one for each judge) in which the number of judges is **an even multiple** of the number of objects. If there are five (5) stimuli then the number of judges must be five (5) 10, 15, 20, etc. The user goes to SAS and selects SOLUTIONS then ANALYSIS then MARKET. When in **Market** the user selects Multidimensional Scaling.

Example with Dessert Data Using SAS Market

In this example 8 students were asked to judge the similarity of four desserts. The desserts are Chocolate Cake, Pound Cake, Chocolate Ice Cream and Vanilla Ice Cream. Note that there are twice as many judges as objects.

A SAS data set is created.

```
data indd;
input judge $ 1 ccake pcake vaice chice dessert $;
datalines;

1 7 5 2 4 ccake
1 5 7 2 2 pcake
1 2 2 7 5 vaice
1 4 2 5 7 chice
2 7 6 2 4 ccake
2 6 7 3 2 pcake
2 3 3 7 7 vaice
2 4 2 7 7 chice
3 7 4 2 4 ccake
3 4 7 4 2 pcake
3 2 4 7 4 vaice
3 4 2 4 7 chice
4 7 6 4 5 ccake
4 6 7 3 3 pcake
4 4 3 7 6 vaice
4 5 3 6 7 chice
5 7 5 3 6 ccake
5 5 7 3 2 pcake
5 3 3 7 5 vaice
5 6 2 5 7 chice
6 7 1 1 3 ccake
6 1 7 5 2 pcake
6 1 5 7 4 vaice
6 3 2 4 7 chice
7 7 4 4 4 ccake
7 4 7 5 2 pcake
7 4 5 7 3 vaice
7 4 2 3 7 chice
8 7 3 1 6 ccake
8 3 7 6 3 pcake
8 1 6 7 4 vaice
8 6 3 4 7 chice

run;
```

This data set is run to produce a WORK file. The WORK file is then handled by the **Market** command or program.

Multidimensional Scaling of WORK.INDD
Analysis Summary
Distance data
Measurement level: Ordinal
objects: CCAKE PCAKE VAICE CHICE
id variable: DESSERT
subject (matrix) variable: JUDGE
number of dimensions: 2 (default)
Stress formula: Kruskal's stress formula 1

WORK file results provide input selections to the analysis.

Based on the WORK file SAS **Market** is entered. The Work file is used and the **Option** to perform Individual Differences analysis is selected. In the analysis summary, Kruskal’s Stress Formula is given in addition to the R^2 fit statistic. As shown in Table 15.1 and Figure 15.1, the relationship between the original distances and the fitted distances is very high.

Table 15.1
Goodness of Fit Statistics for MDS of Several Desserts

No. of Dimen- sions	Judge	Badness of Fit	Distance r	Fit r
2	-	.043	.988	.988
2	1	.006	.999	.999
2	2	.040	.972	.972
2	3	.020	.998	.999
2	4	.070	.922	.922
2	5	.093	.865	.865
2	6	.119	.724	.724
2	7	.024	.998	.998
2	8	.063	.973	.974

Table 15.2
Coordinates of Two Dimensions

Dessert	Dim 1	Dim 2
ccake	0.91	1.02
pcake	1.08	1.07
vaice	-1.00	0.93
chice	-0.99	0.98

The coordinates are plotted in Fig. 15.1. The dimensions are identified as:

- Dimension 1 = Cake vs. Ice Cream.
- Dimension 2 = Chocolate vs. Vanilla

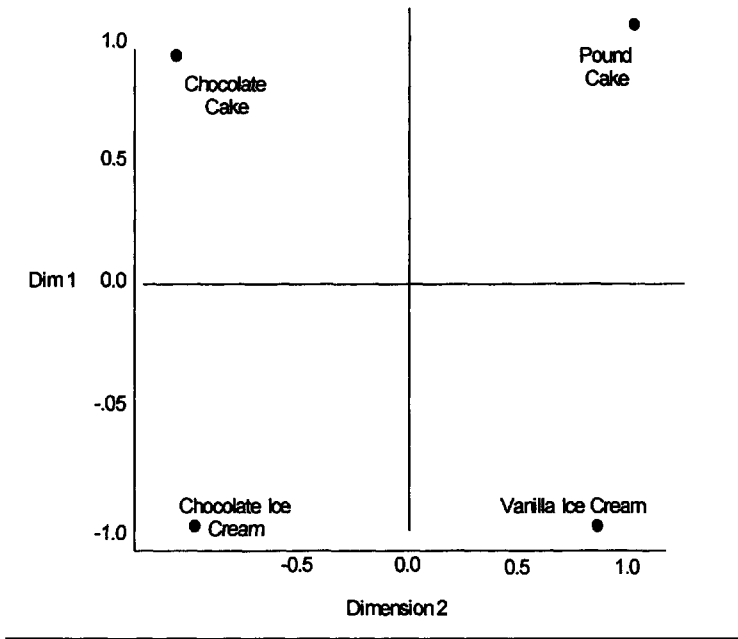


FIG. 15.1 Dessert Configuration.

By checking the **Results** and choosing **coefficients** instead of coordinates the vector positions of each judge are produced as shown in Table 15.3 and in Fig. 15.2.

Table 15.3		
Individual Coefficients of 2 Dimensional Configuration		
Judge	Dim 1	Dim 2
1	1.25	0.66
2	1.32	1.00
3	1.00	0.50
4	1.35	0.43
5	1.07	0.92
6	0.90	1.09
7	0.99	1.01
8	0.48	1.33

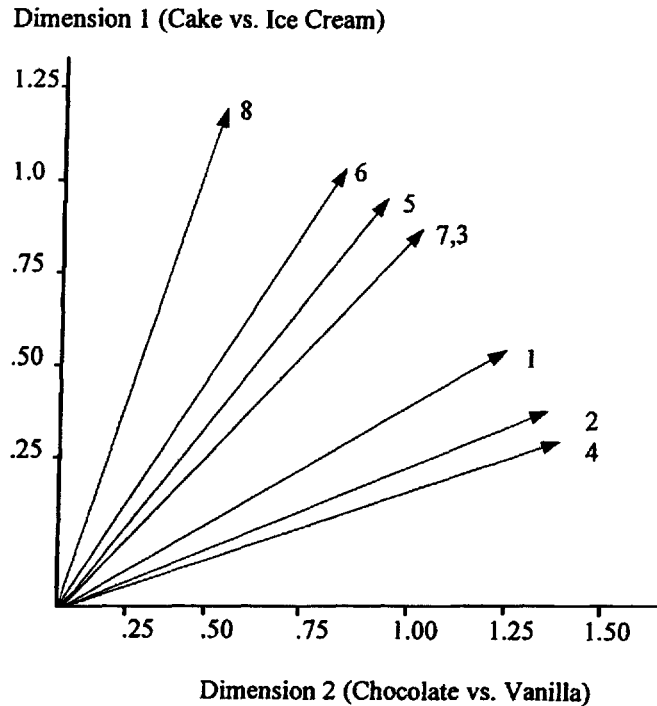


FIG. 15.2. Judge coordinates showing use of dessert dimensions.

Young (Young & Lewycky, 1979) points out that subjects with vectors oriented in the same direction are similar. The end points are not necessarily similar groups. The vector's distance represents the amount of variance accounted for by the two dimensions. In this analysis each subject's variance is effectively determined by one or both of these dimensions. Subject 8 makes decisions primarily based on whether the desserts are icecream or not and subjects 1 and 2 based their judgments on whether the desserts are chocolate or not. Other judges use both dimensions in judging similarity.

How ALSCAL Works

ALSCAL is a repetitive analysis procedure. At each iteration, two steps are taken. These are an **optimal scaling** step and a **model estimation** step. Each of the procedures is a least squares fit and the process alternates back and forth until a criterion is reached (improvement < .001 or number of steps < 30, for example). The steps in the process are:

1. The basic data are usually distances between the stimuli and as a first step an additive constant is calculated so that the triangle inequality holds and there will be no imaginary solutions.
2. A scalar products matrix is computed for each subject and an average B* scalar products matrix is determined. This results in an initial stimulus configuration.
3. Initial weight configuration matrices are computed for each subject.

The optimization algorithm is presented graphically as described by Young and Lewyckyj (1979).

Initial distances in ALSCAL are computed based upon

$$d_{ijk}^2 = \sum w_{ka} (x_{ia} - x_{ja})^2$$

where: w is the weight for subject k on dimension a . x_i and x_j are the coordinates of stimulus i and j on a . The iterative ALSCAL process is diagrammatically displayed in Figure 15.3.

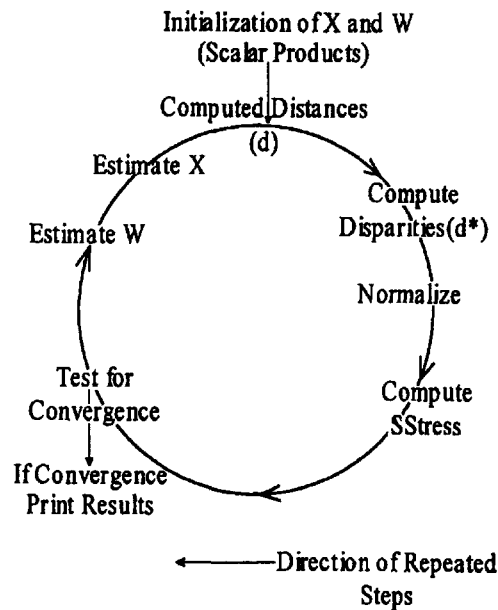


FIG 15.3. Young's convergence diagram.

Alternating Search Analogy

Although the multiple analysis using the alternating least squares approach is not readily demonstrated, the reader can gain some appreciation of the method by observing the simplest technique in an alternating **one dimensional** search. In this example all variables except one are held constant. The remaining one varies. Solutions are obtained until a minimum of the equation or function is reached. The remaining variables are treated similarly one at a time. The process is repeated until a desired solution is obtained.

Suppose, for example, one wished to find the minimum of $Z = 2X^2 + 4Y^2 - 4X$, where X and Y are greater than or equal to zero. First, X is held constant at some arbitrary value. For example $X = 3$ is chosen and the values of Y are varied over the values 3, 2, 1, 0. Solutions for Z are then obtained. The smallest value for Z occurs when $Y = 0$. Next Y is then held constant at zero and X is varied through the values 2, 1, 0. In this case the minimum value of Z occurs when X is

equal to one. Holding X constant at one ($X = 1$), other values of Y are tried to assure that the sixth trial values of $X = 1$ and $Y = 0$ are optimum, that is, result in the minimum value of the function. Figure 15.4 illustrates this solution. The function is $Z = 2X^2 + 4Y^2 - 4X$.

<u>Trials</u>	<u>X</u>	<u>Y</u>	<u>Z</u>
1	3	3	42
2	3	2	22
3	3	1	10
4	3	0	6
5	2	0	0
6	1	0	-2 minimum
7	0	0	0
8	1	1	2
9	1	.5	-1
10	1	.25	-1.75
11	1	0	-2 minimum

FIG. 15.4. An example of an alternating one dimensional search.

The reader should consult Kruskal & Wish (1978) for a more complete explication of INDSCAL. See also Carroll (in Shepard, 1972a; Carroll & Chang, 1968, 1970; Tucker, 1972; and Young, de Leeuw & Takane, 1976).

Application: The Letter Wheel

Some researchers have suggested that the processing of the visual features of letters and words is sequential. But evidence for the hypothesis of integrated or parallel processing has been found. It occurs in an analysis of the reaction times of individual subjects to a set of letters, paired in all possible ways. The length of time it takes for someone to respond to such stimuli (latency) appears to be closely related to reading level. Fifty-two children and adults responded to the question of whether two letters were the same or different (Dunn-Rankin, 1978). In this experiment, two letters placed side by side on a card, are hidden by a shutter. When they are exposed, a clock starts. As soon as the subject presses a switch indicating that the letters are the same or different, the clock stops. The length of time it takes the subject to respond is recorded in hundredths of a second. The reaction time then serves as a measure of letter similarity.

A matrix of these similarities is analyzed using ALSCAL. The 13 letters selected for this study (**f**, **t**, **n**, **h**, **k**, **x**, **z**, **g**, **p**, **q**, **e**, **s**, and **c**) were chosen because they could be combined in various ways to form pairs containing similar letter features. More letters were not included because the increase in paired comparisons makes the task too tiring for young children. The multidimensional scaling analysis takes the form of a circular pattern of letters much like a color wheel. This representation suggests that the dimensions of letters are not immutable but integrative, in other words, one letter melds into the next in a continuous way.

A general division can be suggested for the 13 letters (in terms of angle versus curve or ascender versus descender). The letter **k** is opposite **e**, and **t** is opposite **g**, for example. The letter wheel

indicates that a parallel processing approach to the perception of familiar letters is preferable to a serial model. Most of the 13 letters contain two or more of the basic constituents, and an analogy between primary-color combinations and letter-feature combinations is strongly suggested. Just as orange is seen as a unique color even though it is a combination of red and yellow, the letters of the letter wheel are seen as integrated units that are combinations of basic features. In this model **x** combines with **l** to produce **k**, and **l** combines with **n** to produce **h**, yet **h** and **k** are perceived as wholes.

Further reinforcement for an integrative perception of familiar units comes from the fact that the two basic dimensions indicated (curve versus angle, ascender versus descender) are in general relied on equally by all subjects. When the positions of the 52 subjects with respect to the use they made of these two dimensions were plotted, it was clear that none of the subjects, young or old, good reader or poor, chose one feature exclusively. Some subjects, however, recognized the combination of the two basic features more readily than others. These subjects were also the most mature readers in the sample. Such a result might be an artifact of consistency of response, or might reflect the importance of these dimensions to perception in reading, or both.

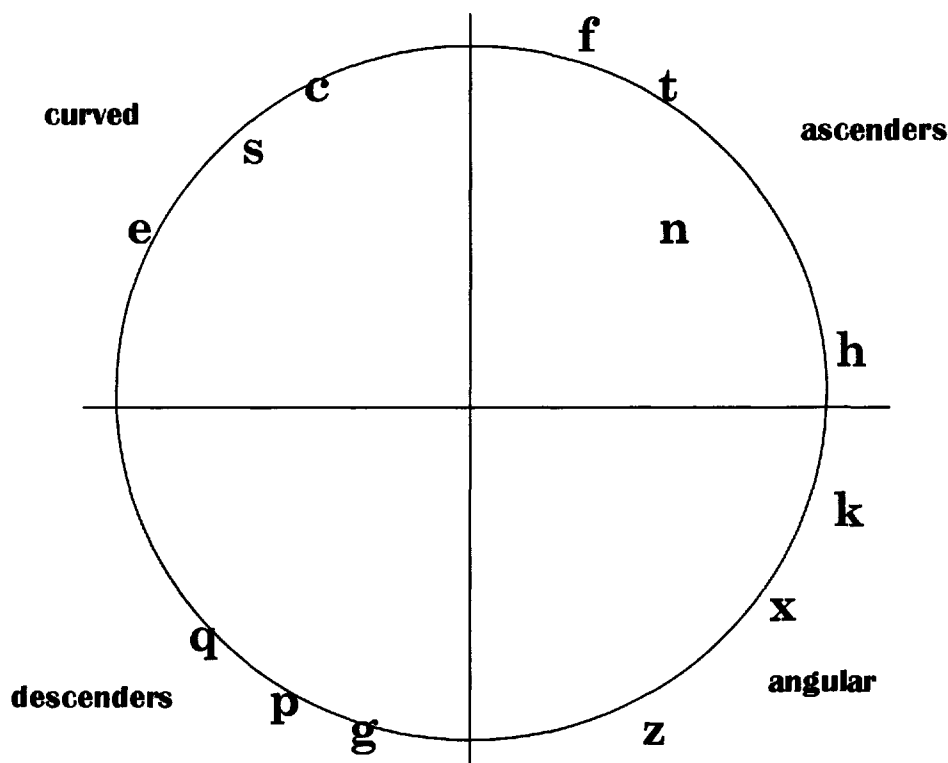


FIG. 15.5. A letter wheel configuration using ALSCAL MDS. Note in this circular representation, extremely unlike letters tend to be opposite each other. The letter **n** is not on the circumference of the wheel because it shares certain characteristics with **e**, **s**, and **c**. Its left vertical component and its striking similarity to **h**, however, place it closer to the letters with vertically ascending components.

This page intentionally left blank

APPENDIX A

Using a Computer to Solve Problems

SAS

Personal Computers are so powerful today that they can handle Scaling programs quite easily. The SAS System is available to teachers and instructors and educational institutions at nominal rates. System Version 8 (V8) can be installed for as little as \$75.00 a year. The authors feel SAS has the most to offer in its data analysis procedures (PROC) and in its Market Research Applications. The **market** command contains two programs (MULTIDIMENSIONAL SCALING, and MULTIDIMENSIONAL PREFERENCE ANALYSIS) which are easy to use and provide functional results. They require that the data be in a particular form. Once the data form has been mastered, the solutions are a matter of selecting options.

Other analyses require a more conventional SAS approach to their solutions. In SAS V8, there is a program editor and work log and an output log. The raw data are written in or copied into the program editor. Here three steps are required: (1) the data definition step, (2) the input step and (3) the raw data. Each statement is ended by a semicolon. For example:

Example of a SAS input file

```
data Measures;  
input height weight sex;  
datalines;  
72 165 1  
74 210 1  
65 123 0  
62 110 0  
;  
proc means;  
run;
```

(Defines the data)
(Variables are separated by spaces)
(Indicates that raw data follows)

(semicolon ends the raw data)
(The means procedures is initiated)
(This last command executes the statistical analysis included in the Means procedure)

This data can be submitted by clicking the run button or writing submit in the command space. The results can be observed and any errors will be marked in the log window.

Format

One format survives from FORTRAN programming that is still functional in defining data. This is the **F** format. Measurements require a decimal point to indicate their value. Suppose one wishes to read a set of 5 numbers such as 24356, each one as a measurement. This would require an **F** or Floating point format. If the data started in column 1, the conventional format is (5F1.0). The 1 after the F indicates how many columns are occupied by each variable. The 5 in front of the F tells how many variables there are. The 1.0 indicates how accurately the data are measured. Here, there are no places to right of the decimal point and the data are read as single digit measures. SAS dispenses with F and 0 and simply uses a number and a decimal point to indicate the degree of measurement. For example (2.)

When data are **not** spaced SAS uses the @ symbol to indicate the column in which the data should be initially read. SAS uses (N1-N10 or more) to indicate the number of variables (N can be any name). Their input statement reads ten values of input data as [Input @ 7 (N1-N10) (2.)].

SAS uses the \$ to indicate that a name should be read. If for example a name was associated with each line of raw scores the input command might be [Input NAME \$ @7 (N1-N10) (2.)].

Delwiche and Slaughter (1998) provided a complete explication for SAS users in their primer while Carey and Carey (1996) provides useful tips.

Using the CD-ROM

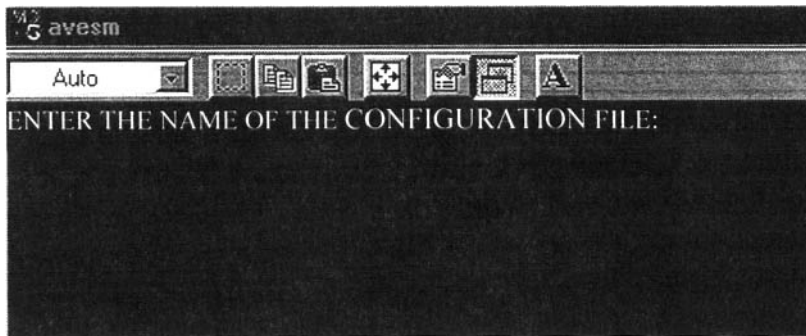
The (CD-ROM) programs do **not** require data in a tightly specified format. All programs use free format data that only requires at least one space between each data entry on a given line. A configuration file is used to convey information about the subjects and objects for each analysis. The first line in this file is the title line. The second line is one in which the parameters are displayed separated by spaces. The third line of the configuration file indicates the file that contains the data (all data values are separated by spaces). The fourth line of this initial file indicates the name of the output file (where the results will be stored).

When a user wants to run a program, for example **AVEMAT**, which averages individual similarity estimates, **avemat.exe** is called. The MS-DOS screen then asks for the configuration file. The user types the name of the file, for example, **avemat.cfg**. The configuration file then provides the title, parameters and the data input and output file names (for example **avemat.dat** and **avemat.out**). The program gets data from the **avemat.dat** and puts the results in **avemat.out**.

The **avemat** folder on the cd contains information about the program **AVEMAT**. It contains;

avemat_readme.txt	(describes the required format of the configuration file, avemat.cfg and the specifications for the data/input file.)
avemat.exe	executable code for Windows based PC's.
avemat.cfg	a configuration file.
avemat.dat	a data or input file.
avemat.out	an output file

Upon executing, `avemat.exe` interactively prompts for the name of the configuration file. The name is limited to 64 characters. It looks as follows:



You type in the name of the configuration file (for example `avemat.cfg`).

The configuration file, `avemat.cfg`, for AVEMAT has the following format:

- First line: title line
- Second line: 3 values separated by white space (blanks or tabs)
 - first value: number of subjects (integer value)
 - second value: number of variables (integer value)
 - third value: maximum similarity (real value, but does not require a decimal point)
- Third line: name of the input/data file (maximum of 64 characters).
- Fourth line: name of the output file (maximum of 64 characters).

avemat.cfg: Sample `avemat` configuration file

```

Avemmat Sample Data
5 4 7
avemat.data
avemat.output
  
```

Specifications for the input data file: (`avemat.dat`)

- First line(s) The pairwise keys, each member of the pair and each pair delimited by white space [tabs, blanks, end of lines (eolns)].
- Subsequent line(s): Pairwise proximity data (usually similarities) for each subject, values delimited by white space (tabs, blanks, eolns). Each subject's data may spread over more than one line **BUT each persons's data must begin on a new line**. Characters on the line after the last value for each subject are ignored and can be used for identification purposes, if desired.

Sample avemat data file [5 subjects, 4 objects (6 pairs)]

01	02	01	03	01	04	02	03	02	04	03	04
5	2	4	2	2	5					DS	
6	2	4	3	2	7					HA	
4	2	4	4	2	4					DB	
6	4	5	3	3	6					BB	
5	3	6	3	2	5					AF	

The same data file with a different form -- pairs are specified with single digits (over two lines) -
- blanks are replaced by tabs -- id information is eliminated -- two lines for each subject's re-
sponses

1	2	1	3	1	4
2	3	2	4	3	4
6	2	4			
2	2	5			
6	2	4			
3	2	7			
4	2	4			
4	2	4			
6	4	5			
3	3	6			
5	3	6			
3	2	5			

Output file created with above data file: (avemat.out).

Avemat Sample Output Data

Variables = 4 Subjects = 5

7.00	5.20	2.60	4.60
5.20	7.00	3.00	2.20
2.60	3.00	7.00	5.40
4.60	2.20	5.40	7.00

Readme General

On the CDROM is a general readme file which is also reported here. It is useful to read the file to avoid mistakes. The computer programs accompanying this book implement many of the techniques described in the text. They are adapted from the programs in the first edition of Scaling Methods to run on a personal computer. These programs do not need data in a fixed format, i. e., they all use "list directed" I/O also called variable format data. Each program (with a .exe suffix) uses a small configuration (or runtime control) file (with a .cfg suffix) which provides necessary information to the program. Each program has an accompanying "readme" file (with

a .txt suffix) that describes information specific to that program and sample data (.dat suffix) and report files (.out suffix). This document describes characteristics that apply to all of the programs. You should read this entire file before using any of the programs.

SYSTEM REQUIREMENTS

These software programs are stand-alone programs. This means that you can use any one of them independently of the others. It also means that you don't need any additional software for them to run. The only software that you need is a text editor, like Notepad or Edit that come with Microsoft Windows.

The computer programs were designed for an Intel Pentium-class PC running Microsoft Windows95 or later. The computer should have at least 32Meg of main memory and at least 3Meg of free space on the hard disk drive. A printer would be useful if you want to print a report or graph and the programs will work with any printer installed in Microsoft Windows.

PREPARING TO RUN THE PROGRAMS

a. Creating Text Files

In order to use your own data with a program you must create two input files: a configuration file and a data file. Each of these should be created with a pure text editor like Notepad or Edit. Word processors, like Microsoft Word or WordPerfect, often include hidden characters that will keep the programs from running correctly, if at all.

When you create a file, you should be certain that the file name does not have a "suffix" (extra characters at the end of the file name) added automatically by the editor. For example, Notepad may add ".txt" without telling you. To complicate matters, Microsoft Windows can hide the file name suffix when you look at the files in a folder. To make sure that you can see the full file name, select View from the window's menu, select Folder Options, select View, and then make sure that "Hide file extensions for known file types" has NOT been selected. If a check mark is in that box, you can clear it by clicking on it. An important note about Notepad: be sure to press the Enter key after you have typed the last line in the file. Otherwise, the last line will not be processed correctly.

b. Configuration Files

Each program will use a "control" file called a configuration file to execute. The configuration file is a small file that contains:

- (1) A title for the report produced.
- (2) Information such as the number of subjects and the number of variables needed by that particular program.
- (3) The name of the data file that contains the raw or input data.
- (4) The name of the output file that will contain the report.

The "readme" file for each program will describe exactly how the information should appear in a configuration file for that program.

All file names (configuration, input and output) may be up to 64 characters long. An example for a configuration file for PEROVER could be named `perover1.cfg` and contain the following lines:

```
Data for the Spring Semester 2002
16 30
Perover.dat
Perover.out
```

The name of the configuration file is one of your choosing. It is good practice to provide configuration files with a consistent extension such as ".cfg" as in the example above. In this case, the configuration file would need to be in the same folder as the executable program you are currently running.

In the above sample configuration file, the input file named `Perover.dat` is also expected to be in the same folder as `perover.exe` and the report file `Perover.out` will be placed in that same folder. If there is no file named `perover.cfg` or no file named `Perover.dat` in the expected folder, there will be an "Input file does not exist" error message along with the name of the missing file.

A report file named `Perover.out` will be created and placed in the same folder as `perover.exe`. The report file will be overwritten every time you run the program without changing the report file name. However, before overwriting an existing file, each program will check to see if the report file exists and, if it does, ask you if you want to replace (overwrite) it. If you want to save the old report, give the old report file a new name or change the name of the new report file in the program's configuration file. If you want to use a configuration file or input file in some other folder or create a report file in some other folder you may do so by using full file names.

c. Data Files

Data files use a "free form" format. This means that the data items are separated by "white space" (spaces, tabs or new lines). The data items do not have to be in specific columns. A consequence of this approach means that the programs do not handle missing data items - there must be a value for every data item. One data set may take more than a single line but each new data set must begin on a new line. The content and order of the data items is described in the "readme" file for each program. It is good practice to provide data files with a consistent extension such as `.dat`.

RUNNING THE PROGRAMS

While it is possible to run the programs from the CD-ROM or a floppy diskette, the programs will be easier to use and will run faster if they are copied to a folder or subdirectory on the hard disk drive and run from there. This will also help if the output is large since floppy diskettes have limited space and reports cannot be written to CD-ROMs.

Once the program has been copied to the hard disk drive, you can start the program from "My Computer" or Windows Explorer. For those whose experience with microcomputers predates

Microsoft Windows, the programs may also be run from an MS-DOS prompt. If you are using "My Computer" or Windows Explorer, double click on the executable program name (the file name will end with .exe, for example perover.exe).

A small (MS-DOS) window will open and you will be asked for the name of the configuration file. If the configuration file is not in the same folder as the executable program, type the full file name. After typing in the configuration file name, press Enter. If the report file already exists, the program will ask you if you want to replace it. (The trcir program also prompts for additional information which is described in the "readme" file for that program.) If the program executes correctly, a message will appear telling you that the program is finished. If the program encounters an error, an error message will be displayed. Remember or write down the error message so that you can fix the problem (see probable causes of errors in section 5). Press the Enter key and the window will disappear. The report from the program will be located in the folder you specified in the configuration file. You may now view the report or print it. It is good practice to use a consistent extension such as .out for report files.

PRINTING REPORTS

The report files created by these programs are simple text files with no more than 80 characters per line. So that columns and graphs line up correctly, they should be printed in a nonproportional font such as Courier or Mishikawa. If the output of one program is, after editing, going to be used as input for another program, it must be saved as a simple text file.

ERROR MESSAGES

Message: Unexpected end of input reading file <file name>

Cause: The program was expecting more data than was found in the file.

Possible solution(s):

- (1) Check the file named in the configuration file to make sure that it contains all of the data.
- (2) Make sure that there are "white spaces" between all of the data items.
- (3) If you created the file named in the error message with Notepad, be sure that you pressed the Enter key after the last value in the file.
- (4) If the file named in the error message is the configuration file, make sure that the items in the file are on the correct lines and in the correct order as described in the program's "readme" file.
- (5) If the file named in the error message is the data file, make sure that the value for the number of subjects (in the configuration file) agrees with the number of data sets in the data file. Also, make sure that each data set has the correct number of values.

Message: Input file does not exist. Name: <file name>

Cause: The program cannot find the file called <file name>.

Possible solution(s):

- (1) If the file named in the error message is the configuration file, make sure that you typed it correctly when the program started executing.

- (2) If you are sure that you typed the configuration file name correctly, make sure that Microsoft Windows has not automatically added a suffix to the file name.
- (3) If the file named in the error message is the data file, make sure that the data file name in the configuration file matches the data file name on disk and that Microsoft Windows has not automatically added a suffix to the file name.

Message: Error while reading input file. Name: <file name>

Cause: The program has tried to read a value that doesn't match the type of value it is expecting. For example, the program is expecting to read digits but the next characters in the file are not digits.

Possible solution(s):

- (1) Make sure that numeric values are really numeric. For example, make sure that an 'O' hasn't been substituted for a '0' (zero) or an 'l' for a '1' (one).
- (2) Make sure that the correct number of values of each type is in the file and that the values are in the order described by the program's "readme" file.
- (3) Make sure that you used a pure text editor instead of a word processor that could put hidden characters in the file.

Message: The number of <variable name> given is greater than the maximum number of <variables> this program can handle.

Cause: The value, read from the configuration file for the variable named in the error message, exceeds the maximum value allowed.

Possible solution(s):

Make sure that the value of <variables> in the configuration file is correct.

TROUBLESHOOTING

Problem: In the report file, *** appear in some places where numeric values should be.

Cause: The computed values are too large to fit in the number of columns allotted for that value.

Solution: Care has been taken to allow enough space for all output values. However, if this problem occurs, check the Scaling Methods website for a different version of the program. If one is not found, contact the authors.

FULL FILE NAMES

You may use full file names (up to 64 characters) for the configuration, input and report files. If you use the full file name of a configuration or input file, then that file does not have to be in the same folder as the executable program you are running. Using the full file name for a report file means that the file can be created in any folder, not just the current folder.

A full file name contains the disk drive, any folder names, and the file name separated by backslashes ('\'). For example, if you wanted a configuration file called "perover.cfg" on the C: drive in a folder called \MyData\Spring02, the full configuration file name would be:

C:\MyData\Spring02\perover.cfg

The data and the reports are identified in the configuration file by the names of the files on the hard disk drive. These names are also limited to 64 characters each. For example, if you have data file called "Perover.out" in a folder on the C: drive called \MyData\Spring02, the full name of the data file would be:

C:\MyData\Spring02\Perover.out

This file name contains 30 characters so it would be valid for these computer programs. If you wanted to save the report in a file called "Trial Report" on the C: drive in a folder called \My reports, the full file name would be:

C:\My reports\Trial Report

This report file name has 26 characters so it is also a valid file name for these programs.

WHAT IS INCLUDED ON THE CD-ROM FOR EACH PROGRAM

The CD-ROM contains the following files for each program:

- <program name>_readme.txt -- brief program description with requirements for the configuration and data files, maximum values allowed, etc.

- <program name>.exe -- executable code for Windows based PC's

- <program name>.cfg -- example configuration file

- <program name>.dat -- example data/input file

- <program name>.out -- report file created using the data given

Some programs have other .cfg, .dat and .out files as well.

Using the Internet

Bell Labs Netlib

Bell Telephone Laboratory was a formidable bastion for Clustering and Scaling. Starting with Shepard's original ideas, the group of Joe Kruskal, Douglas Carroll, Mike Wish, J.J. Chang, Sandra Pruzansky, Steve Johnson and others completed an historic litany of non-metric and metric programs for analyzing group and individual judgments and choices. The group of programs, mostly written in FORTRAN, is available on the net. The reader can go to:

<http://netlib.bell-labs.com/netlib/mds/>

to view the entire set of programs which, after giving proper credit, are essentially in the public domain.

The most useful programs are listed below

kyst2a.f (Kruskal's MDS program)

kyst2a.manual.txt

mdpref.f (Carroll and Chang's Preference Vector Program) is on the CD-ROM

prefmap.f (Carroll and Chang's Point Mapping Program)

sindscal.f (Carroll and Pruzanski's Individual Differences Scaling Program) is on CD-ROM

hiclus.f (Johnson's Non-Metric Clustering Program) is on the CD-ROM

It is a relatively easy matter to utilize this rich resource if one has access to a FORTRAN compiler such as the F77 compiler at the University of Hawaii's computer center. The authors of this text were particularly interested in Carroll's individual differences scaling methods refined by Sandra Pruzansky as SINDSCAL (Simplified Individual Differences Scaling). Obtaining and using the program was completed in the following steps.

1. Using the web browser the program (sindscal.f) was located on the list.
2. It was double clicked by mouse and stored in a local home computer on Microsoft's WordPad.
3. Using WS-ftp the program was moved to the authors UNIX file at the University.
4. Two read and write file specifications needed by UNIX were added to the program and it compiled.
5. Using a data file the program was run and the output saved.

Many of the Bell labs programs have been compiled and are also available free from:

<http://www.newmdsx.com/team.htm>

PC-MDS

Scott Smith, in the Dept of Marketing at Brigham Young University, has taken many of the Bell-Labs programs as well as the old BMD (BioMedical) programs and packaged them for distribution. Among their programs are **CLUSTER**, the Howard Harris large group clustering of subjects using K-means and **Case5** Thurstone's paired comparison methodology.

PC-MDS charges for this service. (<http://marketing.byu.edu/htmlpages/pcmds/pcmds.htm>).

ViSta

Forrest Young (1996) at the University of North Carolina has developed **ViSta** a statistical analysis set of programs that includes many scaling methods. **ViSta** can be downloaded free and is an exceptional visual experience:

www.forrest@visualstats.org

The Three-Mode Company

The Three Mode Company offers a series of programs designed to analyze three way data. Three way data can be scores of subjects on several variables at different times. TRILIN for triple linear data can perform parallel factor analysis (PARAFAC) which is another way of handling individual differences. For more information access:

<http://www.fsw.leidenuniv.nl/~kroonenb/document/programs.htm>

ProGAMMA

ProGamma (now Science Plus Group) has a software catalog that offers a program on Mokken Scaling, MSP 5 for Windows. This program is user friendly and is applicable to building unidimensional scales of test items under item response theory as well as doing Guttman Scaling. Its cost is non-trivial.

<http://www.scienceplus.nl>

Scaling Methods and Terms

The Internet is a valuable resource for learning and understanding the terms and methods associated with Scaling.

There are a number of search engines that are available but one of the best is **Google.com**. Suppose for example that you wish to find the angle, Q where the $\cos Q = .81$. If you write **Scientific Calculator** in the search space of www.google.com, you will receive over 60,000 entries in 0.28 seconds, five of which in the first 30 are on-screen calculators you can use. The best we found were:

Scientific Calculator:

www.scientificcalculator.com, and

CoCalc PRN Scientific Calculator:

www.cohort.com.

Fisher's exact test:

www.matforsk.no/ola/fisher.htm.

Factor Analysis Glossary:

www.siu.edu/~epse1/pohlmann/factglos/.

In addition, many statistics, math and educational psychology departments of universities display lectures and demonstrations of topics that can be accessed. Much information about topics such as Singular Value Decomposition (SVD) and Multidimensional Scaling (MDS) can be found through search engines such as Google on the World Wide Web.

Distributions

The interested reader can download a free program PQRS which details a wide number of distributions. If for example you wished to determine the normal value for any proportion you can simply click on the interactive graph.

www.eco.rug.nl/medewerk/knypstra/pqrs.html

APPENDIX B

Tables

List of Tables

Table A: Balanced Orders for Pair Comparisons for the Numbers from Five to Seventeen

Table B: Selected Balanced Incomplete Block Designs

Table C: Percentage Points of the Studentized Range for Infinite Degrees of Freedom

Table D: Selected Range Values in the Two-Way Classification

Table E: Cumulative Probability Distribution for Circular Triads Upper and Lower 10% Tails
Across 5-15 Objects

Table A
Balanced Orders for Paired Comparisons
For the Numbers from Five to Seventeen

N = 5	9 cont.	11 cont.	13 cont.	13 cont.
1-2	8-6	8-9	7-10	12-1
5-3	7-1	1-4	8-9	7-6
4-1	4-3	3-5	1-3	8-5
3-2	5-2	2-6	2-4	9-4
4-5	6-9	11-7	13-5	10-3
1-3	7-8	10-8	12-6	11-2
2-4	1-4	9-1	11-7	12-13
5-1	3-5	5-4	10-8	1-7
3-4	2-6	6-3	9-1	6-8
<u>2-5</u>	9-7	7-2	4-3	5-9
	8-1	8-11	5-2	4-10
N = 7	5-4	9-10	6-13	3-11
1-2	6-3	1-5	7-12	2-12
7-3	7-2	4-6	8-11	13-1
6-4	8-9	3-7	9-10	7-8
5-1	1-5	2-8	1-4	6-9
3-2	4-6	11-9	3-5	5-10
4-7	3-7	10-1	2-6	4-11
5-6	2-8	6-5	13-7	3-12
1-3	9-1	7-4	12-8	<u>2-13</u>
2-4	5-6	8-3	11-9	
7-5	4-7	9-2	10-1	N = 15
6-1	3-8	10-11	5-4	1-2
4-3	<u>2-9</u>	1-6	6-3	15-3
5-2		5-7	7-2	14-4
6-7	N = 11	4-8	8-13	13-5
1-4	1-2	3-9	9-12	12-6
3-5	11-3	2-10	10-11	11-7
2-6	10-4	11-1	1-5	10-8
7-1	9-5	6-7	4-6	9-1
4-5	8-6	5-8	3-7	3-2
3-6	7-1	4-9	2-8	4-15
<u>2-7</u>	3-2	3-10	13-9	5-14
	4-11	<u>2-11</u>	12-10	6-13
N = 9	5-10		11-1	7-12
1-2	6-9	N = 13	6-5	8-11
9-3	7-8	1-2	7-4	9-10
8-4	1-3	13-3	8-3	1-3
7-5	2-4	12-4	9-2	2-4
6-1	11-5	11-5	10-13	15-5
3-2	10-6	10-6	11-12	14-6
4-9	9-7	9-7	1-6	13-7
5-8	8-1	8-1	5-7	12-8
6-7	4-3	3-2	4-8	11-9
1-3	5-2	4-13	3-9	10-1
2-4	6-11	5-12	2-10	4-3
9-5	7-10	6-11	13-11	5-2

15 cont.	15 cont.	17 cont.	17 cont.	17 cont.
6-15	9-4	10-1	3-7	12-3
7-14	10-3	3-2	2-8	13-2
8-13	11-2	4-17	17-9	14-17
9-12	12-15	5-16	16-10	15-16
10-11	13-14	6-15	15-11	1-8
1-4	1-7	7-14	14-12	7-9
3-5	6-8	8-13	13-1	6-10
2-6	5-9	9-12	6-5	5-11
15-7	4-10	10-11	7-4	4-12
14-8	3-11	1-3	8-3	3-13
13-9	2-12	2-4	9-2	2-14
12-10	15-13	17-5	10-17	17-15
11-1	14-1	16-6	11-16	16-1
5-4	8-7	15-7	12-15	9-8
6-3	9-6	14-8	13-14	10-7
7-2	10-5	13-9	1-6	11-6
8-15	11-4	12-10	5-7	12-5
9-14	12-3	11-1	4-8	13-4
10-13	13-2	4-3	3-9	14-3
11-12	14-15	5-2	2-10	15-2
1-5	1-8	6-17	17-11	16-17
4-6	7-9	7-16	16-12	1-9
3-7	6-10	8-15	15-13	8-10
2-8	5-11	9-14	14-1	7-11
15-9	4-12	10-13	7-6	6-12
14-10	3-13	11-12	8-5	5-13
13-11	2-14	1-4	9-4	4-14
12-1	15-1	3-5	10-3	3-15
6-5	8-9	2-6	11-2	2-16
7-4	7-10	17-7	12-17	17-1
8-3	6-11	16-8	13-16	9-10
9-2	5-12	15-9	14-15	8-11
10-15	4-13	14-10	1-7	7-12
11-14	3-14	13-11	6-8	6-13
12-13	2-15	12-1	5-9	5-14
1-6		5-4	4-10	4-15
5-7	N = 17	6-3	3-11	3-16
4-8	1-2	7-2	2-12	2-17
3-9	17-3	8-17	17-13	
2-10	16-4	9-16	16-14	
15-11	15-5	10-15	15-1	
14-12	14-6	11-14	8-7	
13-1	13-7	12-13	9-6	
7-6	12-8	1-5	10-5	
8-5	11-9	4-6	11-4	

For even numbers of pairs, the next higher odd set is used by striking all pairs containing the nonexistent odd object.

Table B
Selected Balanced Incomplete Block Designs

PLANS

Plan 13.1 $t = 7, k = 3, r = 3, b = 7, L = 1, E = .78$, Type II

Block	Reps.		
	I	II	III
(1)	7	1	3
(2)	1	2	4
(3)	2	3	5
(4)	3	4	6
(5)	4	5	7
(6)	5	6	1
(7)	6	7	2

Plan 13.3 $t = 11, k = 5, r = 5, b = 11, L = 2, E = .88$, Type I

Block	Reps.				
	I	II	III	IV	V
(1)	1	2	3	4	5
(2)	7	1	6	10	3
(3)	9	8	1	6	2
(4)	11	9	7	1	4
(5)	10	11	5	8	1
(6)	8	7	2	3	11
(7)	2	6	4	11	10
(8)	6	3	11	5	9
(9)	3	4	10	9	8
(10)	5	10	9	2	7
(11)	4	5	8	7	6

Plan 13.5 $t = 13, k = 4, r = 4, b = 13, L = 1, E = .81$, Type I

Block	Reps.			
	I	II	III	IV
(1)	13	1	3	9
(2)	1	2	4	10
(3)	2	3	5	11
(4)	3	4	6	12
(5)	4	5	7	13
(6)	5	6	8	1
(7)	6	7	9	2
(8)	7	8	10	3
(9)	8	9	11	4
(10)	9	10	12	5
(11)	10	11	13	6
(12)	11	12	1	7
(13)	12	13	2	8

Plan 13.9 $t = 16, k = 6, r = 6, b = 16, L = 2, E = .89$, Type I

Block	Reps					
	I	II	III	IV	V	VI
(1)	1	2	3	4	5	6
(2)	2	7	8	9	10	1
(3)	3	1	13	7	11	12
(4)	4	8	1	11	14	15
(5)	5	12	14	1	16	9
(6)	6	10	15	13	1	16
(7)	7	14	2	16	15	3
(8)	8	16	12	2	4	13
(9)	9	15	11	5	13	2
(10)	10	11	6	12	2	14
(11)	11	4	16	3	9	10
(12)	12	3	10	15	8	5
(13)	13	6	9	14	3	8
(14)	14	13	5	10	7	4
(15)	15	9	4	6	12	7
(16)	16	5	7	8	6	11

Plan 13.13 $t = 21, k = 5, r = 5, b = 21, L = 1, E = .84$, Type I

Block	Reps.				
	I	II	III	IV	V
(1)	21	1	4	14	16
(2)	1	2	5	15	17
(3)	2	3	6	16	18
(4)	3	4	7	17	19
(5)	4	5	8	18	20
(6)	5	6	9	19	21
(7)	6	7	10	20	1
(8)	7	8	11	21	2
(9)	8	9	12	1	3
(10)	9	10	13	2	4
(11)	10	11	14	3	5
(12)	11	12	15	4	6
(13)	12	13	16	5	7
(14)	13	14	17	6	8
(15)	14	15	18	7	9
(16)	15	16	19	8	10
(17)	16	17	20	9	11
(18)	17	18	21	10	12
(19)	18	19	1	11	13
(20)	19	20	2	12	14
(21)	20	21	3	13	15

Plan 13.14 $t = 31, k = 6, r = 6, b = 31, L = 1, E = .86, \text{Type I}$

Block	Reps.					
	I	II	III	IV	IV	VI
(1)	31	1	3	8	12	18
(2)	1	2	4	9	13	19
(3)	2	3	5	10	14	20
(4)	3	4	6	11	15	21
(5)	4	5	7	12	16	22
(6)	5	6	8	13	17	23
(7)	6	7	9	14	18	24
(8)	7	8	10	15	19	25
(9)	8	9	11	16	20	26
(10)	9	10	12	17	21	27
(11)	10	11	13	18	22	28
(12)	11	12	14	19	23	29
(13)	12	13	15	20	24	30
(14)	13	14	16	21	25,	31
(15)	14	15	17	22	26	1
(16)	15	16	18	23	27	2
(17)	16	17	19	24	28	3
(18)	17	18	20	25	29	4
(19)	18	19	21	26	30	5
(20)	19	20	22	27,	31	6
(21)	20	21	23	28	1	7
(22)	21	22	24	29	2	8
(23)	22	23	25	30	3	9
(24)	23	24	26	31	4	10
(25)	24	25	27	1	5	11
(26)	25	26	28	2	6	12
(27)	26	27	29	3	7	13
(28)	27	28	30	4	8	14
(29)	28	29	31	5	9	15
(30)	29	30	1	6	10	16
(31)	30	31	2	7	11	17

Designs taken from Cochran & Cox, (1957) *Design of Experiments*, with permission from the publisher. John Wiley & Sons, Inc. t = treatments; k = number in each block; r = number of times each object appears in the design; b = number of blocks, L = number of times each object is paired, E = efficiency factor; and Type = type of analysis of variance.

Table C

Percentage Points of the Studentized Range for Infinite Degrees of Freedom

k	.80	.90	.95	.99	.999
2	1.812	2.326	2.772	3.643	4.654
3	2.424	2.902	3.314	4.120	5.063
4	2.784	3.240	3.633	4.403	5.309
5	3.037	3.478	3.858	4.603	5.484
6	3.232	3.661	4.030	4.757	5.619
7	3.389	3.808	4.170	4.882	5.730
8	3.520	3.931	4.286	4.987	5.823
9	3.632	4.037	4.387	5.078	5.903
10	3.730	4.129	4.474	5.157	5.973
11	3.817	4.211	4.552	5.227	6.036
12	3.895	4.285	4.622	5.290	6.092
13	3.966	4.351	4.685	5.348	6.144
14	4.030	4.412	4.743	5.400	6.191
15	4.089	4.468	4.796	5.448	6.234
16	4.144	4.519	4.845	5.493	6.274
17	4.195	4.568	4.891	5.535	6.312
18	4.242	4.612	4.934	5.574	6.347
19	4.287	4.654	4.974	5.611	6.380
20	4.329	4.694	5.012	5.645	6.411
22	4.405	4.767	5.081	5.709	6.469
24	4.475	4.832	5.144	5.766	6.520
26	4.537	4.892	5.201	5.818	6.568
28	4.595	4.947	5.253	5.866	6.611
30	4.648	4.997	5.301	5.911	6.651
32	4.697	5.044	5.346	5.952	6.689
34	4.743	5.087	5.388	5.990	6.723
36	4.786	5.128	5.427	6.026	6.756
38	4.826	5.166	5.463	6.060	6.787
40	4.864	5.202	5.498	6.092	6.816
50	5.026	5.357	5.646	6.228	6.941
60	5.155	5.480	5.764	6.338	7.041
70	5.262	5.582	5.863	6.429	7.124
80	5.353	5.669	5.947	6.507	7.196
90	5.433	5.745	6.020	6.575	7.259
100	5.503	5.812	6.085	6.636	7.314

From Harter et. al. (1959) *Probability Integrals of the Range*. Reprinted with permission.

Table D
Selected Range Values in the Two Way Classification

		Objects: 3-15													
Judges	p	3	4	5	6	7	8	9	10	11	12	13	14	15	
2	.01	0	0	0	0	12	14	16	18	20	22	24	26	28	
	.05	0	0	8	10	12	14	15	17	19	21	23	25	26	
3	.01	0	9	12	14	16	19	22	24	27	29	32	35	37	
	.05	6	8	10	13	15	17	20	22	25	27	30	32	25	
4	.01	8	11	14	17	20	23	26	29	32	35	38	41	45	
	.05	7	10	12	15	18	21	23	26	29	32	35	38	41	
5	.01	9	12	16	19	23	26	29	33	37	40	44	47	51	
	.05	8	11	14	17	20	23	26	30	34	37	40	43	47	
6	.01	10	14	17	21	25	29	33	37	41	45	49	53	57	
	.05	9	12	15	19	22	26	29	33	37	41	43	48	52	
7	.01	11	15	19	23	27	31	36	40	44	49	53	58	62	
	.05	9	13	16	20	24	28	32	36	40	44	48	52	56	
8	.01	12	16	20	25	29	34	38	43	47	52	57	62	67	
	.05	10	14	17	21	25	30	34	38	42	47	51	56	60	
9	.01	12	17	22	26	31	36	41	46	51	56	61	66	71	
	.05	10	14	18	23	27	31	36	40	45	50	54	59	64	
10	.01	13	18	23	28	33	38	43	49	54	59	65	70	75	
	.05	11	15	19	24	28	33	38	43	47	52	57	62	67	
11	.01	14	19	24	29	35	40	46	51	57	62	68	74	78	
	.05	11	15	20	25	30	35	40	45	50	55	60	65	71	
12	.01	14	20	25	31	36	42	48	54	59	65	71	77	83	
	.05	12	16	21	26	31	36	41	47	52	58	63	68	74	
13	.01	15	21	26	32	38	44	50	56	62	68	74	80	87	
	.05	12	17	22	27	32	38	43	49	54	60	65	71	77	
14	.01	16	21	27	33	39	45	52	58	64	71	77	84	90	
	.05	13	17	23	28	34	39	45	50	56	62	68	74	80	
15	.01	16	22	28	34	41	47	54	60	67	73	80	87	94	
	.05	13	18	24	29	35	40	46	52	58	64	70	76	83	

Table E
Cumulative Probability Distributions for Circular Triads Upper and Lower 10% Tails
Across 5-15 Objects

CT	p	CT	p	CT	p
K = 5a		K = 9		K = 11	
0	0.117188	
1	0.234375	4	0.000312	18	0.000539
2	0.468750	5	0.000603	19	0.000827
3	0.703125	6	0.001253	20	0.001275
4	0.976562	7	0.002283	21	0.001889
5	1	8	0.004176	22	0.002846
		9	0.006953	23	0.004110
K = 6		10	0.012213	24	0.005997
0	0.021973	11	0.018673	25	0.008478
1	0.051270	12	0.030023	26	0.012052
2	0.119629	13	0.045212	27	0.016589
...		14	0.067463	28	0.023082
6	0.772949	15	0.095354	29	0.031004
7	0.919434	16	0.137733	30	0.041931
8	1	...		31	0.055270
		25	0.882270	32	0.072952
K = 7		26	0.945247	33	0.093552
0	0.002403	27	0.980161	34	0.120783
1	0.006409	28	0.997580	...	
2	0.016823	29	0.999953	47	0.879692
3	0.032845	30	1	48	0.925622
4	0.068893			49	0.958095
5	0.111992	K = 10		50	0.981092
...		...		51	0.992945
11	0.852943	10	0.000458	52	0.998444
12	0.964294	11	0.000758	53	0.999890
13	0.998741	12	0.001313	54	0.999999
14	1	13	0.002097	55	1
		14	0.003436		
K = 8		15	0.005256	K = 12	
0	0.000150	16	0.008310	...	
1	0.000451	17	0.012131	28	0.000530
2	0.001302	18	0.018354	29	0.000756
3	0.002804	19	0.026093	30	0.001093
4	0.006359	20	0.037960	31	0.001539
5	0.011219	21	0.052003	32	0.002183
6	0.022554	22	0.073128	33	0.003006
7	0.036973	23	0.096711	34	0.004206
8	0.062775	24	0.130882	35	0.005698
9	0.093817	...		36	0.007799
10	0.152757	35	0.888642	37	0.010418
...		36	0.940942	38	0.014012
17	0.858642	37	0.972469	39	0.018372
18	0.949214	38	0.992072	40	0.024359
19	0.987967	39	0.998629	41	0.031372
20	1	40	1	42	0.040744

Table E (continued)

CT	<i>p</i>	CT	<i>p</i>	CT	<i>p</i>
K = 12 (cont.)		K = 13 (cont.)		K = 15	
43	0.051835	85	0.994243	...	
44	0.066124	86	0.997904	76	0.000686
45	0.082403	87	0.999416	77	0.000873
46	0.103535	88	0.999918	78	0.001112
...		89	0.999996	79	0.001406
62	0.892734	90	0.999999	80	0.001778
63	0.928160	91	1	81	0.002232
64	0.957391			82	0.002803
65	0.976516	K = 14		83	0.003495
66	0.989375	...		84	0.004356
67	0.996038	57	0.000650	85	0.005395
68	0.999061	58	0.000859	86	0.006677
69	0.999874	59	0.001121	87	0.008207
70	1	60	0.001467	88	0.010088
		61	0.001897	89	0.012311
K=13		62	0.002460	90	0.015013
...		63	0.003150	91	0.018194
41	0.000582	64	0.004050	92	0.022026
42	0.000801	65	0.005141	93	0.026481
43	0.001087	66	0.006542	94	0.031821
44	0.001478	67	0.008232	95	0.037973
45	0.001977	68	0.010382	96	0.045252
46	0.002656	69	0.012936	97	0.053593
47	0.003511	70	0.016164	98	0.063365
48	0.004653	71	0.019959	99	0.074406
49	0.006081	72	0.024684	100	0.087266
50	0.007957	73	0.030201	101	0.101642
52	0.010265	74	0.036993	...	
52	0.013277	75	0.044800	124	0.890410
53	0.016916	76	0.054337	125	0.915958
54	0.021585	77	0.065172	126	0.937877
55	0.027196	78	0.078191	127	0.955623
56	0.034255	79	0.092853	128	0.969851
57	0.042577	80	0.110238	129	0.980443
58	0.052989	...		130	0.988202
59	0.065036	100	0.887220	131	0.993359
60	0.079797	101	0.916113	132	0.996645
61	0.096719	102	0.940796	133	0.998495
62	0.117057	103	0.959876	134	0.999440
...		104	0.974773	135	0.999833
79	0.871163	105	0.985089	136	0.999966
80	0.907402	106	0.992166	137	0.999996
81	0.936236	107	0.996283	138	0.999999
82	0.959543	108	0.998552	139	0.999999
83	0.975984	109	0.999553	140	1
84	0.987417	110	0.999911		
		111	0.999990		
		112	1		

REFERENCES

- Anderberg, M. R. (1973). *Cluster analysis for applications*. New York: Academic Press.
- Anderson, R. E. (1966). A computer program for Guttman scaling with the Goodenough technique. *Behavioral Science*, 1(3), 235.
- Barr, A. J., Goodnight, J. H., Sall, J. P., & Helwig, J. (1976) *A user's guide to SAS 761*. Raleigh, NC: SAS Institute, Inc.
- Bashaw, W. L., & Anderson, H. E., Jr. (1968). Developmental study of the meaning of adverbial modifiers. *Journal of Educational Psychology*, 59, 111-118.
- Berg, S. R. (1995). *Dynamic Scaling: An ipsative procedure using techniques from computer adaptive testing*. Doctoral Dissertation, Department of Educational Psychology, University of Hawaii, Honolulu.
- Blashfield, R. K., & Oldenderfer, M. S. (1978). The literature on cluster analysis. *Multivariate Behavioral Research*, 13, 271-295.
- Blumenfeld, W. S. (1972, May). I am never startled by a fish. *Industrial Psychologist Newsletter*.
- Boorman, S. A., & Arabie, P. (1972). Structural measures and the method of sorting. In R. N. Shepard & S. B. Nerlove (Eds.), *Multidimensional Scaling* (Vol. 1). New York: Seminar Press.
- Bouma, A. (1971). Visual recognition of isolated lower-case letters. *Vision Research*, 11, 459-474.
- Carey, H., & Carey, G. (1996). *SAS today! A year of terrific tips*. Cary, NC: SAS Institute Inc., p. 395.
- Carroll, J. D. (1972). Individual differences and multidimensional scaling. In R. N. Shepard, A. K. Romney, & S. B. Nerlove (Eds.), *Multidimensional scaling* (Vol. 1). New York: Seminar Press.
- Carroll, J. D., & Arabie, P. (1980). Multidimensional scaling. *Annual Review of Psychology*, 31, 607-649.
- Carroll, J. D., & Chang, J. J. (1968). *Program INDSCAL*. Murray Hill, NJ: Bell Telephone Laboratories.
- Carroll, J. D., & Chang, J. J. (1970). Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart-Young" decomposition. *Psychometrika*, 35, 283-319.
- Cattell, R. B. (1952). *Factor Analysis: An introduction and manual for the psychologist and social scientist*. New York: Harper & Row.
- Cattell, R. B. (1962). The basis of recognition and interpretation of factors. *Educational and Psychological Measurement*, 22, 667-697.

- Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, 1, 245-276.
- Chambers, T. M., & Kleiner, B. (1980). *Graphical techniques for multivariate data and clustering*. Paper presented at the June Annual Meeting of the Classification Society, Boulder, Colorado.
- Chang, J. J. (1968). *Preference mapping program*. Murray Hill, NJ: Bell Telephone Laboratories.
- Cliff, N. (1973). Scaling. *Annual Review of Psychology*, 24, 473-506.
- Cliff, N. (1977). A theory of consistency of ordering generalizable to tailored testing. *Psychometrika*, 42(3), 375-399.
- Cliff, N., & Caruso, J. C. (1998). The factor structure of the WAIS-R: Replicability across agegroups. *Multivariate Behavioral Research*, 33(2), 273-293.
- Cochran, W. G., & Cox, G. M. (1957). *Experimental designs*. New York: Wiley.
- Coombs, H. C. (1964). *A Theory of Data*. New York: Wiley.
- Coombs, C. H. (1967). Thurstone's measurement of social values revisited forty years later. *Journal of Personality and Social Psychology*, 6, 87.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297-334.
- Cudeck, R. (1980). A comparative study of indices for internal consistency. *Journal of Educational Measurement*, 17(2), 117-130.
- David, H. A. (1959). Tournaments and paired comparisons. *Biometrika*, 46, 139-149.
- David, H. A. (1963). *The method of paired comparisons*. New York: Hafner Publishing Company.
- Davison, M. L. (1983). *Multidimensional scaling*. New York: Wiley.
- de Jong, A., & Molenaar, I. W. (1987). An application of Mokken's model for stochastic cumulative scaling in psychiatric research. *Journal of Psychiatric Research*, 21, 137-149.
- Delwiche, L. D., & Slaughter, S. J. (1998). *The little SAS book: A primer (2nd ed)*. Cary, NC: SAS Publishing.
- Dixon, W. J., & Brown, M. B. (1979). *BMDP-79: Biomedical Computer Program P Series*. Berkeley, CA: University of California Press.
- Dixon, W. J., & Massey, F. J. (1969). *Introduction to statistical analysis*. New York: McGraw Hill.
- Donovan, M. A. (1977). *The relationship between modality preferences and programs used in initial reading instruction*. Doctoral dissertation, University of Hawaii, Honolulu.
- Dunn-Rankin, P. (1965). *The true distribution of the range of rank totals and its application to psychological scaling*. Doctoral dissertation, Florida State University, Tallahassee.
- Dunn-Rankin, P. (1966). An IBM 7040 Fortran IV program for constructing scales from paired comparisons. *Behavioral Science*, 110, 234.
- Dunn-Rankin, P. (1968). The similarity of lowercase letters of the English alphabet. *Journal of Verbal Learning and Verbal Behavior*, 7, 990-995.
- Dunn-Rankin, P. (1976, July). *Results of research on the visual characteristics of words*. A paper presented to the Far West Regional Conference of International Reading.
- Dunn-Rankin, P. (1978). The visual characteristics of words. *Scientific American*. January, 238, (1), 122-130.
- Dunn-Rankin, P. (1983). *Scaling methods*. Hillsdale, NJ: Erlbaum.
- Dunn-Rankin, P. (1990). Eye movement research on letters and words. In R. Groner, G.

- d'Ydewalle & R. Parham (Eds.), *From Eye To Mind*. North Holland: Elsevier Science Publishers.
- Dunn-Rankin, P. (1995, April). *A multidimensional look at word similarity*. Paper presented at the 1995 American Educational Research Association Meeting, San Francisco.
- Dunn-Rankin, P. A. (1987). *Abilities and performance in vocabulary acquisition*. (Doctoral Dissertation, University of Hawaii, Honolulu).
- Dunn-Rankin, P., & King, F. J. (1969). Multiple comparisons in a simplified rank method of scaling. *Educational and Psychological Measurement*, Summer, 29(2), 315-329.
- Dunn-Rankin, P., Knezek, G. A., & Abalos, J. (1978, May). *A Circular triads revisited*. Paper presented at the Hawaii Psychological Association meeting, Honolulu, Hawaii.
- Dunn-Rankin, P., Leton, D. A., & Sato, M. (1972). The similarity of hiragana characters in typos 35 font. *The Science of Reading*, 16(2).
- Dunn-Rankin, P., Shimizu, M., & King, F. J. (1969). Reward preference patterns in elementary school children. *International Journal of Educational Sciences*, 31(1), 53-62.
- Dunn-Rankin, P., & Wilcoxon, F. (1966). The true distribution of the range of rank totals in the two-way classification. *Psychometrika*, 3(4), 573-580.
- Dunn-Rankin, P., & Wong, E. (1980, April). *Interjudge similarity following free clustering*. Paper presented to the American Educational Research Association Meeting, Montreal.
- Dunn-Rankin, P., & Zhang, S. (1998). Scaling methods. In J. Keeves (Ed.), *Methodology and measurement: An international handbook* (2nd Ed.) Oxford, UK: Elsevier Science Publishers.
- Eckhart, C., & Young, G. (1936). The approximation of one matrix by another of lower rank. *Psychometrika*, 1(3), 211-218.
- Edwards, A. L. (1957) *Techniques of attitude scale construction*. New York: Appleton Century-Crofts.
- Edwards, A. L. (1959). *Edwards personal preference schedule manual*. New York: The Psychological Corporation.
- Ekman, G. A (1963). Direct method for multidimensional ratio scaling. *Psychometrika*, 28(1), March.
- Everitt, B. (1974). *Cluster analysis*. London: Heinemann.
- Finney, D. J. (1948). The Fisher-Yates test of significance in 2 x 2 contingency tables. *Biometrika*, 35, 145-156.
- Fruchter, B. (1954). *Introduction to factor analysis*. New York: Van Nostrand.
- Furlong, M. J., Atkinson, D. R., & Janoff, D. S. (1980) Elementary school counselors perceptions of their actual and ideal roles. *Elementary School Guidance and Counseling Journal*.
- Gelphman, J. L., Knezek, G. A., & Horn, C. E. (1992). The topological panorama camera: A new tool for teaching concepts related to space and time. *Journal of Computers in Mathematics and Science Teaching*, 11(1), 19-29.
- Gnanadesikan, R. (1977). *Statistical data analysis of multivariate observations*. New York: Wiley.
- Gnedenko, B. V., & Khinchin, A. Y. (1962). *An Elementary Introduction to the Theory of Probability* (Leo F. Born, trans.). New York: Dover. (Original work published 1960)
- Goodenough, W. H. (1944). A technique for scale analysis. *Educational and Psychological Measurement*, 4, 179-190.
- Gorsuch, R. L. (1983). *Factor analysis*. Hillsdale: Lawrence Erlbaum Associates.

- Gower, J. C. (1971) A general coefficient of similarity and some of its properties. *Biometrics*, 27, 857-872.
- Green, B. F. (1954). Attitude measurement. In G. Lindley (Ed.), *Handbook of Social Psychology*. Reading, MA: Addison-Wesley.
- Green, P. E., & Carmone, F. J. (1970). *Multidimensional scaling and related techniques in marketing analysis*. Boston: Allyn & Bacon.
- Guilford, J. P. (1954). *Psychometric methods* (2nd ed.). New York: McGraw-Hill.
- Gulliksen, H. (1958). An IBM 650 program for a complete paired comparisons schedule (Parcoplet 2-21). Tech. Rep ONR Contract Nonr. 1859(15).
- Gulliksen, H., & Tucker, L. R. (1961). A general procedure for obtaining paired comparisons from multiple rank orders. *Psychometrika*, 26, 173-184.
- Gulliksen, H., & Tukey, J. W. (1958). Reliability for the law of comparative judgment. *Psychometrika*, 23(2).
- Guttman, L. (1944). A basis for scaling qualitative data. *American Sociological Review*, 9, 139-150.
- Guttman, L. (1950). The basis for scalogram analysis. In S. A. Stouffer (Ed.), *Measurement and Prediction*. Princeton, NJ: Princeton University Press.
- Harman, H. H. (1967). *Modern factor analysis*. Chicago: University of Chicago Press.
- Harshman, R. A. (1978). *Models for analysis of asymmetrical relationships among n objects or stimuli*. Presented at first joint meeting of Psychonomic Society and Society for Mathematical Psychology, Hamilton, Ontario.
- Harter, L. H. (1959). *The probability integrals of the range and of the studentized range*. WADC Technical Report 58-484, Wright Patterson Air Force Base.
- Hays, W. L. (1973). *Statistics for the social sciences*. (2nd ed.). New York: Holt, Rinehart, and Winston.
- Hiraki, K. (1974). *Teacher status in Japan*. Honors thesis, College of Education, University of Hawaii.
- Horst, P. (1965). *Factor analysis of data matrices*. New York: Holt Rinehart and Winston.
- Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24, 417-441.
- Jacoby, W. G. (1991). *Data Theory and Dimensional Analysis*. Beverly Hills: Sage.
- Johnson, S. C. (1967). Hierarchical clustering schemes. *Psychometrika*, 32, 241-254.
- Kaiser, H. F. (1958) The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, 23, 187-200.
- Kendall, M. G. (1952). *The Advanced Theory of Statistics, Vol. I* (5th ed.). London: Charles Griffin.
- Kendall, M. G. (1955). Further contributions to the theory of paired comparisons. *Biometrics*, 11, 43-62.
- Kendall, M. G., & Babington-Smith, B. (1939). On the method of paired comparisons. *Biometrika*, 31, 324-345.
- Kerlinger, F. N., & Pedhazur, E. J. (1973). *Multiple regression in behavioral research*. New York: Holt, Rinehart, & Winston.
- King, F. J. (1974, August). *A content referenced interpretive system for standardized reading tests*. A final report to the Research Foundation of the National Council of Teachers of English.
- Knezek, G. A. (1979). *Circular triad distributions with applications to complete paired comparisons data*. Doctoral dissertation, University of Hawaii.

- Knezek, G., Wallace, S., & Dunn-Rankin, P. (1998). Accuracy of Kendall's chi-square Approximation to circular triad distributions. *Psychometrika*, 63, 23-34.
- Krtishnaiah (Ed.) *Multivariate Analysis* (Vol. 2), New York: Academic Press.
- Krus, D. J. (1978). Logical basis of dimensionality. *Applied Psychological Measurement*, 37, 587-601.
- Krus, D. J., Bart, W. M., Airasian, P. W. (1975). *Ordering Theory and Methods*, Los Angeles, CA: Theta Press.
- Kruskal, J. B. (1964a) Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, 29, 1-27.
- Kruskal, J. B. (1964b) Nonmetric multidimensional scaling: A numerical method. *Psychometrika*, 29(2), 115.
- Kruskal, J. B., & Carroll, J. D. (1969). Geometric models and badness-of-fit functions. In P. R. Krishnaiah (Ed.), *Multivariate Analysis*. New York: Academic Press.
- Kruskal, J. B., & Wish, M. (1978). *Multidimensional scaling*. Beverly Hills: Sage.
- Kuder, G. F., & Richardson, M. W. (1937). The theory of the estimation of test reliability. *Psychometrika*, 2(3), 151-160.
- Kuenappas, T., & Janson, A. J. (1969). Multidimensional similarity of letters. *Perception and Motor Skills*, 28, 3-12.
- Levy, S., & Guttman, L. (1975). On the multivariate structure of well-being. *Social Indicators Research*, 2, 361-388.
- Likert, R. A. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, 140, 5-53.
- Loevinger, J. (1948). The technique of homogeneous tests compared with some aspects of "scale analysis" and factor analysis. *Psychological Bulletin*, 45, 507-529.
- Mahoney, T. (1986). Seriousness of crimes: Then and now. Unpublished Masters Paper, Department of Educational Psychology, University of Hawaii, Honolulu.
- Marascuilo, L. A., & McSweeney, M. (1977). *Nonparametric and distribution-free methods for the social sciences*. Monterey, CA: Brooks/Cole.
- McClarty, J. (1980). Personal communication.
- McQuitty, L. (1957). Elementary linkage analysis for isolating orthogonal and oblique types and typal relevancies. *Educational and Psychological Measurement*, 17, 207-229.
- McRae, D. J. (1971). MIKCA, A Fortran IV iterative k means cluster analysis program. *Behavioral Science*, 16, 423-434.
- Mokken, R. J., & Lewis, C. (1982). A nonparametric approach to the analysis of dichotomous item responses. *Applied Psychological Measurement*, 6, 417-430.
- Montenegro, X. P. (1978). *Ideal and actual student perceptions of college instructors as predictors of teacher effectiveness*. Doctoral dissertation, University of Hawaii, Honolulu.
- Moore, D. S. (1994). *The basic practice of statistics*. New York: Freeman & Co.
- Moseley, R. L. (1966). *An analysis of decision making in the controllership process*. Doctoral dissertation, University of Washington, Seattle.
- Mosteller, F. (1951). Remarks on the methods of paired comparisons: A test of significances for paired comparisons when equal standard deviations and equal correlations are assumed. *Psychometrika*, 16, 3-9.
- Mosteller, F. (1958). The mystery of the missing corpus. *Psychometrika*, 23(4).

- Napier, D. (1972). Nonmetric multidimensional techniques for summated ratings. In R. N. Shepard, A. K. Romney, & S. B. Nerlove (Eds.), *Multidimensional scaling* (Vol. 1). New York: Seminar Press.
- Nie, N. H., Hull, C. H., Jenkins, J., Steinbrenner, K., & Bent, D. H. (1975). *SPSS: Statistical package for the social sciences* (2nd ed.). New York: McGraw Hill.
- Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1957). *The measurement of meaning*. Urbana: University of Illinois Press.
- Pang, C. M. (1996). *Using familiarity to order a large lexicon*. Doctoral dissertation, University of Hawaii, Honolulu.
- Pedhazur, E. J., & Schmelkin, L. P. (1991). *Measurement, design and analysis*. Hillsdale: Lawrence Erlbaum Associates.
- Pruzansky, S. (1975). *How to use SINDSCAL: A computer program for individual differences in multidimensional scaling*. Murray Hill, NJ: Bell Telephone Labs.
- Remmers, H. H. (1963). Rating methods in research on teaching. In Gage, N. L. (ed.), *Handbook of Research on Teaching*. Chicago: Rand McNally.
- Robinson, J. P., Rusk, J. G., & Head, K. B. (1969a). *Measures of occupational attitudes*. Ann Arbor ISR.
- Robinson, J. P., Rusk, J. G., & Head, K. B. (1969b). *Measures of political attitudes*. Ann Arbor ISR.
- Robinson, J. P., Rusk, J. G., & Head, K. B. (1969c). *Measures of social psychological attitudes*. Ann Arbor ISR.
- Romney, A. K., Shepard, R. N., & Nerlove, S. B. (1972). *Multidimensional scaling* (Vol. 2). New York: Seminar Press.
- Roskam, E. (1970). Method of triads for nonmetric multidimensional scaling. *Psychologie*, 25, 404-417.
- Ross, R. T. (1934). Optimal orders in the method of paired comparisons. *Journal of Experimental Psychology*, 25, 414-424.
- Rummel, J. F. (1964). *An introduction to research procedures in education*. New York: Harper & Row.
- Rummel, R. J. (1970). *Applied factor analysis*. Evanston, IL: Northwestern University Press.
- Shaw, E. M., & Wright, J. M. (1967). *Scales for the measurement of attitudes*. New York: McGraw-Hill.
- Shepard, R. N. (1962). The analysis of proximities: Multidimensional scaling with an unknown distance function. *Psychometrika*, 27, 125-140.
- Shepard, R. N. (1972a). Introduction to Volume 1. In R. N. Shepard, A. K. Romney, & S. B. Nerlove (Eds.), *Multidimensional scaling* (Vol. 1). New York: Seminar Press.
- Shepard, R. N. (1972b). A taxonomy of some principal types of data and of multidimensional methods for their analysis. In R. N. Shepard, A. K. Romney, & S. B. Nerlove (Eds.), *Multidimensional scaling* (Vol. 1). New York: Seminar Press.
- Siegel, S. (1956). *Nonparametric Statistics for the Behavior Sciences*. New York: McGraw Hill.
- Smith, C. P. (1968). *The distribution of the absolute average discrepancy and its use in significance tests of paired comparison scaling*. Unpublished Master's thesis, University of Hawaii.
- Smith, D. M. (1971). *Another scaling of arithmetic tests*. Unpublished paper, Florida State University, Tallahassee.

- Spath, H. (1980). *Cluster analysis algorithms*. Chichester, UK: Ellis Harwood.
- Starks, T. H. (1958). *Tests of significance for experiments involving paired comparisons*. Doctoral dissertation, Virginia Polytechnic Institute, Blacksburg.
- Starks, T. H., & David, H. A. (1961). Significance tests for paired comparison experiments. *Biometrika*, 48 (1 & 2), 95.
- Subkoviak, M. J. (1975). The use of multidimensional scaling in educational research. *Review of Educational Research*, 45(3), 387-423.
- Swartz, R. (2002). *CT3 as an index of knowledge domain structure: Distribution for order analysis and information hierarchies*. Doctoral dissertation, University of North Texas, Denton.
- Takane, Y., Young, F., & de Leeuw, J. (1976) Nonmetric individual differences multidimensional scaling: An alternating least squares method with optimal scaling features. *Psychometrika*, 42, 7-67.
- Thurstone, L. L. (1927). The method of paired comparisons for social values. *Journal of Abnormal and Social Psychology*, 21, 384-385.
- Thurstone, L. L. (1947). *Multiple factor analysis*. Chicago: University of Chicago Press.
- Thurstone, L. L. (1927). A law of comparative judgment. *Psychological Review*, 34, 251-259.
- Torgerson, W. S. (1958). *Theory and methods of scaling*. New York: Wiley.
- Tucker, L. R. (1972). Relations between multidimensional scaling and three mode factor analysis. *Psychometrika*, 37, 3-27.
- Veldman, D. J. (1967) *Fortran programming for the behavioral sciences*. New York: Holt, Rinehart, & Winston.
- Villanueva, M., & Dunn-Rankin, P. (1973, April). *A comparison of ranking and rating methods by multidimensional matching*. A paper presented to the American Educational Research Association.
- Waern, Y. (1972). Graphic similarity analysis. *Scandinavian Journal of Psychology*.
- Ward, J. H. (1963) Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, 58, 236-244.
- Ward, J. H. (1980). Personal communication.
- White, D. R. (1998). Statistical entailment analysis 2.0 user's manual. Available <http://eclectic.ss.uci.edu/~drwhite/entail/emanual.html>, Retrieved 12/29/2003.
- Wilcoxon, F., & Wilcox, R. A. (1964). *Some rapid approximate statistical procedures*. New York: Lederle Laboratories.
- Wise, S. L., & Tatsuoka, M. M. (1986). Assessing the dimensionality of Dichotomous data using a modified order analysis. *Educational and Psychological Measurement*, 46, 295-301.
- Wold, H. (1966) Estimation of principal components and related models by iterative least squares. In P. R. Krishnaiak (Ed.), *Multivariate analysis*. New York: Academic Press.
- Young, F. W. (1981). Quantitative analysis of qualitative data. *Psychometrika* 46, 357-388.
- Young, F. W. (1985). Multidimensional scaling. In Kotx-Johnson (Ed.) *Encyclopedia of Statistical Sciences*, Vol. 5, New York: Wiley.
- Young, F. W. (1996). *ViSta: The Visual Statistics System*. Chapel Hill, NC: L.L. Thurstone Psychometric Laboratory.
- Young, F. W., de Leeuw, J., & Takane, Y. (1976). Regression with qualitative and quantitative variables: an alternating least squares method with optimal scaling features. *Psychometrika*, 41, 505-29.

This page intentionally left blank

AUTHOR INDEX

A

Abalos, J., 68
Airasian, P. W., 83
Anderberg, M. R., 124
Atkinson, D. R., 60, 67

B

Babington-Smith, B., 66, 67, 71
Bart, W. M., 83
Berg, S.R., 53
Blashfield, R. K., 143
Blumenfield, W. S., 106
Bouma, A., 116

C

Carey, G., 200
Carey, H., 200
Carmone, F. J., 34
Carroll, J. D., 161, 162, 164, 174, 185, 190,
196, 208
Caruso, J. C., 159
Cattell, R. B., 158
Chambers, T. M., 145
Chang, J. J., 161, 164, 174, 185, 190, 196, 208
Cliff, N., 89, 159
Cochran, W. G., 12, 216
Coombs, H. C., 97
Cox, G. M., 12, 216
Cronbach, L. J., 89, 106, 111
Cudeck, R., 89

D

Davison, M. L., 175
de Jong, 82
de Leeuw, J., 196
Delwiche, L.D., 200
Dixon, W. J., 27, 58, 59
Donovan, M., 141
Dunn-Rankin, P., xi, 12, 15, 24, 27, 33, 55, 57,
59, 62, 63, 66, 68, 97, 103, 116, 117,
124, 143, 196

E

Edwards, A. L., 71, 76,
Ekart, 161
Ekman, G. A., 32

F

Finney, D. J., 88
Fisher, R. A., 83, 88, 89
Fruchter, B., 153
Furlong, M. J., 60

G

Gelphman, J. L., 90
Gnedenko, B. V., 67, 68, 90
Goodenough, W. H., 67, 76, 80, 81
Gower, J. C., 49
Green, B. F., 19, 100
Green, P. E., 34
Guilford, J. P., 60, 153

Gulliksen, H., xii, 12, 97
Guttman, L., 9, 60, 75, 80, 108

H

Harman, H. H., 155, 158
Harshman, R. A., 32
Harter, L. H., 58, 59, 217
Hays, W. L., 57
Head, K.B., 111
Hiraki, K., 170
Horn, C. E., 90

J

Janoff, D. S., 60
Janson, A. J., 116
Johnson, S. C., 132, 134, 135, 208

K

Kaiser, H. F., 158
Kendall, M., 40, 66, 67, 71
Khinchin, A. Y., 67, 68
Kleiner, B., 145
King, F. J., 55, 59, 79, 124
Knezek, G. A., 12, 60, 66, 67, 68, 90
Kruskal, J. B., 175, 178, 196, 208
Krus, D. J., 83
Kuennapas, T., 116

L

Levy, 108
Lewis, 81
Lewyckyj, R., 190, 194, 195
Likert, R., 18, 105
Loevinger, 89

M

Mahoney, T., 97
Massey, F. J., 27, 58, 59
McRae, D. J., 141
McClarty, J., 108
McQuitty, L., 118
Mokken, R. J., 81
Molenaar, 82

Moore, 115
Mosteller, F., 60, 97

O

Oldenderfer, M. S., 143
Osgood, C. E., 19

P

Pang, C. M., 53
Pedhazur, E. J., 152
Pruzansky, S., 208

R

Remmers, H. H., 106
Robinson, J. P., 111
Ross, R. T., 15, 33, 97
Rummel, R. J., 60, 155
Rusk, J. G., 111

S

Schmelkin, L. P., 152
Shaw, F. M., 111
Shepard, R. N., 11, 175, 196
Shimizu, M., 124
Siegel, S., 74
Slaughter, S. J., 200
Smith, C. P., 60, 98
Smith, D. M., 80
Suci, G. J., 19
Swartz, R., xiii, 90

T

Tannenbaum, P. H., 19
Takane, Y., 196
Thurstone, L. L., xii, 15, 60, 93, 97
Torgerson, W. S., 31
Tucker, L. R., xii, 12, 196
Tukey, J. W., 97

V

Veldman, D. J., 19, 103
Villanueva, O., 33

W

Waern, Y., 115
Wallace, S., xi, 12, 66
Ward, J. H., 121, 124
White, D., 91
Wish, M., 178, 196, 208
Wilcox, R. A., 57
Wilcoxon, F., 57
Wong, E., 24
Wright, J. M., 111

Y

Young, F. W., 161, 175, 190, 194, 195, 196,
209

Z

Zhang, S., xii, 21

This page intentionally left blank

SUBJECT INDEX

A

Activities

- example of pairing, 12, 55
- preference scale, 57

Adjectives

- example of pairing, 68, 69
- preference scale, 69

Agglomerative clustering, 121

Alpha, Cronbach's, 111

Application

- Variance Stable Rank Sums, 55
- Circular Triad Analysis, 69
- Guttman Scaling, 75, 79, 80
- Mapping Individual Preference, 161
- Multidimensional Scaling, 175

ALSCAL, 190-197

Association, measures of, 37-52

Attitude statements, construction rules for, 99

Attitude toward reading scale, 4, 99

Attitudinal measurement described 3, 4, 9

AVEMAT (averages matrices), description of, 34-35

Axes, orthogonal, 159

B

Balanced incomplete block design, 15

- example of, 212-216

BIB (paired data from incomplete blocks), 13, 16-17

Binomial test, 71

C

Case V unidimensional scaling, 93, 94

- reliability, 97

Categorical ratings, 18, 99-111

- introduced, 18

ORDER analysis of, 75-92

- sample run, 87

SCALO (Guttman analysis of), 76-81

- sample run, 81

TSCALE analysis of, 19, 103

- sample run, 104

Categories, agreement, 110

Categorizing, 21, 117

Centroid, 30, 137-139

Cheybshev's inequality, 67

Choices, 5

Circularity

- example of, 74

- overall, 67, 70

- pairwise, 70

Circular Triads, 66

Clustering, 113

- free, 23, 25

- k means iterative, 137

- hierarchical-divisive, 143

- instrument, example of, 23

- nonmetric, 132

- partitioning, 137

- pairing and quantifying, 30

- procedures in, 113

- steps in, 113

when to stop, 131

Clusters

number of, 145

graphing of, 145, 146

Cloze tests in reading, 79

Coefficient of reproducibility, 77, 79, 106, 109

scalability, 107

variation, 72

COMPPC (paired comparisons scaling), 97

Correlation, 37

gamma, 40, 42

Kendall's tau, 40-42

Pearson's, 37-40

significance of, 39

squared, 40

squared multiple, 192

Cronbach's Alpha, 111, 152

Cross products, 166

D

Dendogram, 123, 133, 135, 146

types of, 146

Desserts, ranking of, 163

DISSIM (distances), 42

Differences between judges, 24, 49

Dissimilarity, 37

Distance, 31, 42

city block, 43

SAS macro for, 52

Euclidean, 8, 42, 43

formula, 42

psychological, 3, 8, 57

Mahalanobis, 43

Minkowski, 43

Distribution

uniform, 57

normal, 93

Divisive Clustering, 143

E

Eigenvalues, 154

Error Variance, 153

Euclidean Space, 8, 30, 175

Evaluation of instruction, form for, 18, 21

Examples for

clustering, 113-146

multidimensional scaling, 175-179

unidimensional scaling, 55-111

F

Factor

analysis, 149-159

coefficients, 154

loadings, 150

matrix, 150

PC analysis, 155

Factors, number of, 158

Foreign language attitude scale, 108

Free Clustering

introduced, 11

techniques, 23-24

PEROVER (percent overlap)

description of, 23, 47-48

sample run, 26

JUDGED (interjudge distance)

sample run, 27

G

Gamma, Goodman-Kruskal, 42

Goodenough's method, 76-77

GOWER (similarity index), 49- 51

Graph of letter similarity, 8, 63, 116, 123, 197

Graphic similarity analysis, 115

example of, 8, 117

Guttman scaling illustrated, 75

H

HICLUS (nonmetric clustering), 134, 135

I

Ideal point, 174

projection of, 174

Index of rank scalability, 57

Individual differences, 185

Individual differences scaling, 185-198

Interjudge differences

distribution of, 24

example of, 24, 49

mean and variance, 24

Intransitivity, 66

Inversions, 14, 106, 113

Iteration, 176

J

Judge circular triads (JCT), 66
JUDGED (clustering judges), 25, 27
Judges, 57-59
Judgments, 9, 99

K

Kendall's tau correlation, 40-42
KENTAU (SAS Kendall's tau), 41

L

Latency, 33
Letter similarity graph of, 123
 scales, 63
 wheel, 197
Likert scale, 18
Likert scaling, 105-106
Likert scoring, 19
Linkage analysis, 118

M

Mahalanobis d squared, 43
Matrix
 factor, 150
 object by object, 23
 of similarities, 51
 of rank differences, 58
 rank of, 150
 transpose, 150-152
 triangular, 23

Metrics

Euclidean, 8, 42, 43
city block, 43
Minkowski, 43

MDPREF (preference scaling), 15, 161-174
Minimal marginal reproducibility, 79
Minkowski metric, 43

O

Object circular triads (OCT), 67
Object scalability, 70
ORDER, computer program, 87
Ordered categories, 86
Ordered category scaling, steps in, 20

Ordered category ratings, 22

Ordering

BIB program for block designs, 12,
 13, 15, 17
Guttman's Scale (SCALO program),
 9, 75-82
MDPREF analysis of, 15, 147, 164,
 174
pairwise, 11
partial direct, 12-13
RANKO program for ranks, 15, 55-65
TRICIR program for circular triads,
 15, 66-74

P

Pairing and quantifying, 11-13
Pairs, arrangement of, 15
Pairs of pairs, 14
Partitioning, 137
Pearson's r, 22, 37-40
Percentage improvement, 79
Percent overlap, 24, 47-48
Perfect scale, 9
PEROVER (percent overlap), 24, 26, 51
Placing objects, 15, 24, 26
Plot (MDS), of words, 182
Preference, 6, 28
Preference for rewards, 56
Preference mapping, 161-174
PREFMAP program, 174
Principal components factor analysis, 155-158
Profile, 15
Proportions as normal deviates, 94
Proportions, cumulative, 95-97
Proximities, measures of, 37-52
Proximity, 37
 to centroid, 137-139

Q

Qualitative data, scaling, 49-51
Quantifying objects, 17-23
Quantitative data, 8

R

Range of rank totals, distribution of, 57, 218
Ranking

- conditional, 32
- direct, 14, 33
- instrument, example of, 12, 62
- pairs, 29
- pairwise, 12
- partial, 12-13
- tasks, versus rating, 33, 34
- Rank values (RV), 56
- r correlation, 37
- RANKO (rank scaling), 15, 64
 - sample run, 64
- Ratio estimation, 32
- Reading
 - attitude scale, 99, 103
 - modality preference in, 141, 142
- Reflection, 106, 113
- Reference axes, 174
- Relationships, analysis of, 15
- Reliability
 - Cronbach's alpha, 89, 111
 - in the Case V model, 97
- Reproducibility
 - coefficient of, 78
 - minimum marginal, 79
- Reward Preference
 - scales, 124-126
 - profiles, 126, 127
- Residual correlation matrix, 152
- Rotation, 157, 159
- RSQ, 130
- S**
- Sample size, 58, 59
- SAS software, 38, 41, 52, 155, 190
- Scalar product, 45, 164
- Scalability
 - coefficient of, 78
 - index in rank scaling, 57, 58
- SCALAR (scalar products), 45, 164
- Scale
 - construction, steps, 20
 - of adjectives, 69, 73
 - of counselor roles, 60, 61
 - of letters, 63
 - of reading attitude, 99, 103
 - of school subject attitude, 107

- of foreign language attitude, 108, 110
- of reward preference, 65, 124
- score, 55
- type, 11
- value, 56, 103, 110
- Scaling
 - description, 3-6
 - individual differences, 185-197
 - iterative nature, 9
 - multidimensional, 9
 - rank sum, 55-65
 - unidimensional, 53-111
- SCALO program for Guttman Scaling, 76-81
- Scalogram analysis, 78-80
- Scoring unidimensional scales, 19, 55
- Scree test for the number of factors, 158
- Semantic differential, 19
- Similarity Indices
 - Kendall's tau correlation, 40-42
 - Gamma, 42
 - Gower's, 49-51
 - Pearson's r, 37-40
- Similarity Judgments
 - AVEMAT, program for averaging, 28, 34
 - sample run, 35
 - between pairs of objects, 14, 27-29
 - converted to distance, 42
 - direct estimation, 47
 - Gower's measure of, 49
 - INDMAT program for individual
 - matrices, 28, 34
 - sample run, 36
 - matrix output, 185
- SINDSCAL (individual differences), 34, 185
 - sample run, 186-189
- Spatial representation (mapping), 4
- Specific variance, 153
- SPSS software, 111
- Squared multiple correlation, 152
- Standard score, 94, 97
- Subject vectors, 164, 189
- Successive categories, 18
- Successive interval scaling, 100-103
- Summated ratings, 18
- Sums of squares, 140

T

- Tasks, types of, 4
- Tau, Kendall's, 40-42
- Tetrads, 14
- Thurstone's Case V, application of, 97-98
- Total variance, 153
- Transpose matrix, 152
- Triadic comparisons, 30
- Triangle inequality, 44
- TRICIR (circular triads), 71
 - sample run, 72-74
- TSCALE program for successive intervals
 - discussion of, 103
 - sample run, 104

U

- Uniqueness, 153

V

- Variance, in factor analysis, 153
- Variation, coefficient of, 68
- Vector, 165, 189, 194

W

- Word similarity
 - dendogram of, 146

Z

- Z score, 70, 86, 101

This page intentionally left blank

