

## MEASUREMENT, SCALING, AND DIMENSIONAL ANALYSIS

**Course Objectives:** Consider the three terms that are combined in the title of this course: “Measurement” is an operation that is fundamental to scientific research; however, its implications and consequences are often poorly understood. Greater appreciation for the nature of measurement is useful for discerning and exploiting systematic structure in empirical data. Next, the terms “scaling” and “dimensional analysis” refer to a wide variety of research strategies and procedures. The common element among them is that they all seek to provide quantitative and/or geometric representations of the internal structure in a set of data. Researchers apply these techniques for three main reasons: (1) Simple data reduction—summarizing a large set of variables with a smaller number of composite measures; (2) examining dimensionality—testing the underlying sources of variation in a dataset; and (3) measurement—obtaining empirical representations of the underlying (and usually unobservable) dimensions, which can be employed as analytic variables in other statistical procedures. On a less formal note, researchers will often find that dimensional analysis is very beneficial for conceptualizing the contents of their data. In addition, these techniques usually provide visual displays that are very useful for presenting analytical results to other people. Thus, for a variety of reasons, scaling and dimensional analysis are useful additions to the social scientist’s “repertoire” of research strategies.

**Course Prerequisites:** This course assumes that students are familiar, and comfortable, with basic descriptive statistics and the multiple regression model. Some prior exposure to matrix algebra would also be helpful, but it is not absolutely required. However, we will encounter certain mathematical operations which are undefined outside the context of matrices (i.e., the singular value decomposition). Therefore, anyone who has never worked with matrix algebra should be sure to attend the ICPSR Summer Program course, “Mathematics for Social Scientists, II.”

**Course Requirements:** Regular attendance and active class participation is expected. This is an essential component of the course: Statistical knowledge is cumulative, and gaps in the early material will always have detrimental consequences later on. Homework assignments will be given frequently (usually, one assignment every three days or so). A few of these will be problems requiring pencil-and-paper calculations. But, most of the assignments will be computer-based data analysis exercises. All of them are intended to familiarize you with the various concepts and techniques introduced in class and in the readings. Assignments will not be graded for correct answers. But, they will be checked for completion, and comments will be provided. For those participants who are taking this workshop for formal course credit, or for a formal grade letter, a very brief research paper is also required. In this paper, participants are expected to apply one or more of the techniques covered in class to a substantive topic in their own areas of specialization. The details of this paper will be discussed in class. For this subset of course participants, grades will be determined as follows:

|     |                      |
|-----|----------------------|
| 25% | Class attendance     |
| 50% | Homework assignments |
| 35% | Research paper       |

**Software Considerations:** With very few exceptions, the methods covered in this workshop are computationally intensive. Therefore, appropriate software is required to perform most of the analyses. We will rely primarily upon STATA and (for those who are interested) the R statistical computing environment. Note that the other major statistical packages (e.g., SAS, SPSS, etc.) also contain routines that will carry out most of the techniques that we will cover in this course. From time to time, we may use some of these other packages, as well as some special-purpose software, to carry out certain analyses.

**Reading Material:** Unfortunately, there is no single textbook that covers all of the topics in this course. In addition, many of the texts that are available have certain drawbacks that limit their usefulness for our purposes: They tend to be very expensive; they usually assume a high level of mathematical sophistication; they often contain sections that are out of date. Because of these considerations, we will rely primarily on several shorter works taken from the Sage series on Quantitative Applications in the Social Science (i.e., the “little green books”):

Arabie, Phipps; J. Douglas Carroll; Wayne S. DeSarbo (1987) *Three-Way Scaling and Clustering*.

Dunteman, George H. (1989) *Principal Components Analysis*.

Jacoby, William G. (1991) *Data Theory and Dimensional Analysis*.

Kim, Jae-On and Charles W. Mueller (1978a) *Introduction to Factor Analysis*.

Kim, Jae-On and Charles W. Mueller (1978b) *Factor Analysis: Statistical Methods and Practical Issues*.

Kruskal, Joseph B. and Myron Wish (1978) *Multidimensional Scaling*.

McIver, John and Edward G. Carmines (1981) *Unidimensional Scaling*.

Weller, Susan C. and A. Kimball Romney (1990) *Metric Scaling: Correspondence Analysis*.

The following textbooks are also quite good. Some students may want to supplement or replace the Sage Papers with entries from this list:

Bartholomew, David J.; Fiona Steele; Irimi Moustaki; Jane I. Galbraith. (2008) *Analysis of Multivariate Social Science Data (Second Edition)*. CRC Press.

Borg, Ingwer and Patrick Groenen (2005) *Modern Multidimensional Scaling: Theory and Applications (Second Edition)*. Springer.

Cudek, Robert and Robert C. MacCallum, Editors (2007) *Factor Analysis at 100*. Lawrence Erlbaum.

Gorsuch, Richard L. (1983) *Factor Analysis (Second Edition)*. Lawrence Erlbaum.

Greenacre, Michael. (2007) *Correspondence Analysis in Practice (Second Edition)*. Chapman and Hall.

Lattin, James; J. Douglas Carroll; Paul E. Green (2003) *Analyzing Multivariate Data*. Brooks/ColeThomson Learning.

Wickens, Thomas D. (1995) *The Geometry of Multivariate Statistics*. Lawrence Erlbaum.

**Course Web Site:** The home page for this course is located at the following URL:

<http://www.polisci.msu.edu/jacoby/icpsr/scaling>

The contents of this website will evolve and expand as the course proceeds through the subject matter. You should regard the site as an information resource. It will contain the syllabus, copies of handouts, datasets, assignments, computing and software resources, lecture outlines, and links to other interesting and useful sites on the Worldwide Web.

## TOPICS AND READING ASSIGNMENTS

### I. Introductory Concepts

#### A. Measurement Theory and Its Implications

Jacoby (1991), Chapters 1 and 2

Jacoby, William G. (1999) "Levels of Measurement and Political Research: An Optimistic View." *American Journal of Political Science* 43: 271-301.

Weller and Romney (1990), Chapter 1.

Lattin et al. (2003), Chapter 1.

Young, Forrest W. (1981) "Quantitative Analysis of Qualitative Data." *Psychometrika* 46: 357-388.

De Veaux, Richard D. (1990) Finding Transformations for Regression Using the ACE Algorithm." In John Fox and J. Scott Long (Editors), *Modern Methods of Data Analysis*. Sage.

#### B. Data Theory

Jacoby (1991), Chapter 3.

Young, F.W. (1984) "Scaling." *Annual Review of Psychology* 35: 55-60.

#### C. Dimensionality

Jacoby (1991), Chapter 4.

Weisberg, Herbert F. (1974) "Dimensionland: An Excursion into Spaces." *American Journal of Political Science* 18: 743-776.

### II. Exploiting Metric Information in Data

#### A. The Unidimensional Unfolding Model

McIver and Carmines (1981), pp. 71-86.

Coombs, Clyde H. (1950) "Psychological Scaling Without a Unit of Measurement." *Psychological Review* 57: 145-158.

Long, John Frederick and Paul H. Wilken. (1974) "A Fully Nonmetric Unfolding Technique: Interval Values from Ordinal Data." In H.M. Blalock (Editor) *Measurement in the Social Sciences: Theories and Strategies*. Aldine.

Poole, Keith T. (1984) "Least Squares Metric, Unidimensional Unfolding." *Psychometrika* 49: 311-323.

#### B. The Summated Rating Model

Jacoby (1991), pp. 38-41.

McIver and Carmines (1981), pp. 22-40.

Spector, Paul E. (1992) *Summated Rating Scale Construction*. Sage University Paper.

Greene, V.L. and E.G. Carmines (1979) "Assessing the Reliability of Linear Composites." In Karl Schuessler (Editor) *Sociological Methodology 1980*. Jossey-Bass.

Sijtsma, Klaas (2009) "On the Use, Misuse, and the Very Limited Usefulness of Cronbach's Alpha." *Psychometrika* 74: 107-120. Also see the responses and commentaries that follow this article.

### III. Preparation for Multidimensional Models

#### A. Vector Geometry and Linear Models

Wickens (1995), Chapters 1-5.

Lattin et al. (2003), pp. 19-32.

#### B. Singular Value Decomposition and the Basic Structure of a Matrix

Weller and Romney (1990), Chapter 2.

Lattin et al. (2003), pp. 32-36.

Borg and Groenen (2005), pp. 146-163.

### IV. Dimension Reduction and Summarizing Multivariate Data

#### A. The Biplot

Gabriel, K.R. (1971) "The Biplot Graphic Display of Matrices with Application to Principal Components Analysis." *Biometrics* 58: 453-467.

Jacoby, William G. (1998) *Statistical Graphics for Visualizing Multivariate Data*. Sage. Chapter 7.

Gower, J. C. and D. J. Hand (1996) *Biplots*. Chapman and Hall.

## B. Principal Components Analysis

- Dunteman (1989), Chapters 1-6, 8.
- Weller and Romney (1990), Chapter 3.
- Bartholomew et al. (2008), Chapter 5.
- Lattin et al. (2003), Chapter 4.
- Borg and Groenen, pages 519-526.

## V. Factor Analysis

### A. The Common Factor Model

- Jacoby (1991), pp. 47-53.
- Wickens (1995), Chapter 9.
- Kim and Mueller (1978a), pp. 1-46.
- Gorsuch (1983), Chapters 1-4.
- Bartholomew et al. (2008), pages 175-183.
- Lattin et al. (2003), pp. 127-131.
- Bartholomew, David J. (2007) "Three Faces of Factor Analysis." Chapter 2 in Cudek and MacCallum.

### B. Estimating the Factor Model

- Kim and Mueller (1978a), pp. 46-70.
- Kim and Mueller (1978b), pp. 7-29, 41-46.
- Gorsuch (1983), Chapters 6 and 8.
- Bartholomew et al. (2008), pages 183-188.
- Lattin et al. (2003), pp. 131-153
- Jöreskog, Karl G. (2007) "Factor Analysis and Its Extensions." Chapter 5 in Cudek and MacCallum.
- MacCallum, Robert C.; Michael W. Browne; Li Cai. (2007) "Factor Analysis Models as Approximations." Chapter 9 in Cudek and MacCallum.

### C. Rotation

- Kim and Mueller (1978b), pp. 29-41.
- Gorsuch (1983), Chapters 9-10.
- Bartholomew et al. (2008), pages 188-192.
- Lattin et al. (2003), pp. 153-156.
- Jennrich, Robert I. (2007) "Rotation Methods, Algorithms, and Standard Errors." Chapter 14 in Cudek and MacCallum.

D. Constructing Factor Scales

Kim and Mueller (1978b), pp. 60-73.

Gorsuch (1983), Chapter 12.

Bartholomew et al. (2008), pages 192-207.

Lattin et al. (2003), pp. 156-166.

Widaman, Keith F. (2007) "Common Factors Versus Components: Principals and Principles, Errors and Misconceptions. Chapter 10 in Cudek and MaCallum.

E. Introduction to Confirmatory Factor Analysis (if time permits)

Kim and Mueller (1978b), pp. 46-60.

Gorsuch (1983), Chapter 7.

Bartholomew et al. (2008), pp. 289-301.

Lattin et al. (2003), Chapter 6.

Brown, Timothy A. (2006) *Confirmatory Factor Analysis for Applied Research*. Guilford Press.

**VI. Multidimensional Scaling**

A. Spatial Distance Models

Jacoby (1991), pp. 58-62.

Kruskal and Wish (1978), pp. 1-19.

Borg and Groenen (2003), Chapters 1, 17-19.

Lattin et al. (2003), pp. 206-211.

B. Classical Multidimensional Scaling (Metric and Nonmetric)

Kruskal and Wish (1978), pp. 19-60.

Bartholomew et al. (2008), Chapter 3.

Lattin et al. (2003), pp. 211-235.

Borg and Groenen (2005), Chapters 2, 3, 8, 9, 11-13.

Buja, Andreas; Deborah F. Swayne; Michael L. Littman; Nathaniel Dean; Heike Hofmann; Lisha Chen. (2008) "Data Visualization with Multidimensional Scaling." *Journal of Computational and Graphical Statistics* 17: 444-472.

C. Weighted Multidimensional Scaling

Arabie, Carroll, and DeSarbo (1987), pp. 1-53.

Kruskal and Wish (1978), pp. 60-73.

Borg and Groenen (2005), Chapter 22.

Lattin et al. (2003), pp. 235-243.

D. Data for Multidimensional Scaling Analyses

Kruskal and Wish (1978), pp. 73-82.

Jacoby (1991), pp. 63-66.

Borg and Groenen (2005), Chapter 6.

VII. **Multidimensional Unfolding and Preference Mapping**

Jacoby (1991), pp. 66-70.

Weller and Romney, PP. 44-54.

Lattin et al. (2003), pp. 244-252.

Borg and Groenen (2005), Chapters 14-16.

Carroll, J.D. (1972) "Individual Differences and Multidimensional Scaling." In Roger N. Shepard, A. Kimball Romney, Sara Beth Nerlove (Editors), *Multidimensional Scaling: Theory and Applications in the Behavioral Sciences (Volume I, Theory)*. Seminar Press: pp. 105-155.

VIII. **Correspondence Analysis**

Weller and Romney (1990), Chapters 5-8.

Bartholomew et al. (2008), Chapter 4.

Borg and Groenen (2005), Chapter 24.

Greenacre (2007).

IX. **Some Final Considerations: Comparison of Scaling Strategies**

Jacoby (1991), pp. 72-73, Chapter 7.

MacCallum, R.C. (1974) "Relations Between Factor Analysis and Multidimensional Scaling." *Psychological Bulletin* 81: 505-516.

Van Schuur, Wijbrandt H. and Henk A. L. Kiers (1994) "Why Factor Analysis is Often the Incorrect Model for Analyzing Bipolar Concepts and What Model to Use Instead." *Applied Psychological Measurement* 18: 97-110.