

CORRESPONDENCE ANALYSIS OF VOTING DATA

JAN DE LEEUW AND PAULETTE LLOYD

1. INTRODUCTION

Suppose we have a finite domain O of *objects*, and m *categorical variables* ϕ_j . All m variables ϕ_j have the same finite range \mathcal{K} . Thus

$$\phi_j : O \rightarrow \mathcal{K}.$$

Data can be collected in an $n \times m$ matrix (or data-frame) X , where $x_{ij} \in \mathcal{K}$ is defined by $x_{ij} = \phi_j(o_i)$. Binary data are a familiar special case, with $\mathcal{K} = \{0, 1\}$. We use K for $\mathbf{card}(\mathcal{K})$.

In this paper we will concentrate on the following example. There are n individuals voting on m issues. In this example, the K categories will usually be “aye”, “nay”, “abstain”, “absent”, and so on. But another obvious example, with the same formal structure, are *rating scale data*, in which a number of individuals rate a number of things in one of K ordered categories such as “agree”, “indifferent”, “disagree”, “completely disagree”. With our main example in mind, we call data of this type *voting data*.

Date: January 27, 2005.

2000 Mathematics Subject Classification. 62H25.

Key words and phrases. Multivariate Analysis, Correspondence Analysis.

2. HOMOGENEITY

Each variable ϕ_j defines a partition of the n objects into the K subsets

$$O_{j\ell} = \{o \in O \mid \phi_j(o) = \mathcal{K}_\ell\}$$

Clearly there are mK such subsets. A homogeneity analysis finds a representation $\xi : O \rightarrow \mathbb{R}^p$ of all objects in low-dimensional Euclidean space \mathbb{R}^p in such a way that the representations $\xi(O_{j\ell})$ of the mK subsets are relatively small and compact. In other words, points corresponding to objects in the same subset should be relatively close to each other. Or, more specifically, the within-subset variation must be small relative to the total variation.

If we make this precise and quantitative, by using a measure of homogeneity, we can actually quantify how homogeneous a particular low-dimensional representation of the objects is. And we can construct algorithms to find the most homogeneous representation that is possible.

If we only have a single variable we can always attain perfect homogeneity by mapping each of the K subsets of the partition into a single point. If there are more variables, which will be the usual case, then trivial solutions are avoided by requiring that all variables simultaneously have representations which are as homogeneous as possible, or, more specifically, that the average homogeneity over variables is maximized.

There are many possible ways to define the size of a point set in \mathbb{R}^p . Some of them are discussed in De Leeuw [2003]. We mention the diameter, the radius of the smallest circle containing all points, the circumference of the convex hull, the length of the minimum spanning tree, the sum of all distances, or the sum of all distances to the Weber point. We shall use another

measure of homogeneity, which generally leads to much simpler computations.

In *multiple correspondence analysis* the size σ of a point set with t points is defined as the sum of the t squared distances between the points and their centroid. This is closely related to $\tilde{\sigma}$, the sum of squares of all $\binom{t}{2}$ distances between the points. In fact $\tilde{\sigma} = t\sigma$. Geometrically we can picture homogeneity of a point set by adding the lines connecting the points with their centroid, thus forming what we shall call a *star plot*. Homogeneous categories form small compact stars, homogeneous variables form K small stars that are relatively far apart. The star plot of a variable is just the overlay of the K star plots of its categories.

2.1. Loss. Suppose G_j is the indicator matrix of variable j . Thus G_j is a binary matrix, with n rows and K columns, which has exactly one non-zero element in each row. In fact $(G_j)_{i\ell} = 1$ if object i maps into the ℓ^{th} category of variable j , Column ℓ of G_j contains the indicator function of the subset $O_{j\ell}$.

If X is an $n \times p$ matrix with representations of the n objects in \mathbb{R}^p . We only consider matrices X that are centered, i.e. have column sums equal to zero. All reasonable definitions of homogeneity are invariant under translations.

The K centroids of the categories are given by $Y_j = (G_j'G_j)^{-1}G_j'X$. The sum of squared distances of the object points to their categories is

$$\begin{aligned}\sigma_j(X) &= \mathbf{tr} (X - G_j Y_j)'(X - G_j Y_j) = \\ &= \mathbf{tr} X' Q_j X,\end{aligned}$$

where $Q_j = I - P_j$ and $P_j = G_j(G_j'G_j)^{-1}G_j'$. Observe that $X'Q_jX$ is the within-category covariance matrix of X and P_j is the between-category covariance matrix.

Homogeneity requires the within-category covariances to be relatively small, which we interpret as being small compared to the total variance $X'X$. This means we only consider matrices X such that $X'X = I$, and we minimize the average within-category variance over such matrices. This is equivalent to maximizing the between-category variance. Making the stars small is the same, for a given total variance, as making the stars as far apart as possible.

Thus multiple correspondence analysis chooses a centered X with $X'X = I$ such that $\text{tr } X'P_\star X$ is as large as possible, where P_\star is the average of the m matrices P_j . Computationally this means we choose X equal to the p largest eigenvectors of the doubly-centered matrix

$$\tilde{P}_\star = P_\star - \frac{1}{n}uu'$$

where u is a vector with all elements equal to +1. The maximum homogeneity is equal to the sum of the p corresponding eigenvalues.

2.2. Graph Layout. We can think of the $n \times mK$ indicator super-matrix $G = (G_1 \mid \cdots \mid G_m)$ as the adjacency matrix of a bipartite graph with $n + mK$ edges and with nm vertices. Graph layout makes a picture of the graph in \mathbb{R}^p in such a way that a criterion measuring the quality of the layout of the points corresponding with edges and the lines corresponding with vertices is optimized. Quality of the layout can be measured in many different ways, but one obvious one in our context is

$$\sigma(X, Y) = \frac{1}{m} \sum_{j=1}^m \text{tr}(X - G_j Y_j)'(X - G_j Y_j).$$

This is the sum of squares of the lengths of all lines in the layout. And it is exactly the loss function minimized by multiple correspondence analysis if we impose the normalization condition $X'X = I$.

2.3. Incomplete Indicators.

3. VOTING EXAMPLE

Now suppose we specialize this to a voting example where n voters vote “aye” or “nay” or abstain on m resolutions. Thus $K = 3$ and each resolution divides the voters into three sets.

Before we compute the solution, we have to make a number of choices. One option is to only make the “aye” votes homogeneous. Thus we want the sets of points corresponding with the “aye” votes to be small, and we do not care about the sets corresponding with “nay” votes or abstains. The “aye” points could form a little star in \mathbb{R}^p and the “nay” points could be all around it. The second option could be to make both “aye” and “nay” homogeneous, which means that both the “aye” and the “nay” points are ideally in disjoint little stars. And in a third options we could also make the abstain category homogeneous, and try to represent abstain points as little stars as well.

3.1. Transposing.

APPENDIX A. ADA ROLL CALL DESCRIPTIONS

The votes selected cover a full spectrum of domestic, foreign, economic, military, environmental and social issues. We tried to select votes which display sharp liberal/conservative contrasts. In many instances we have chosen

procedural votes: amendments, motions to table, or votes on rules for debate. Often these votes reveal true attitudes frequently obscured in the final votes.

Code	Description	ADA
	Ashcroft Attorney General Confirmation	no
SJ Res 6	Ergonomics Rule Disapproval	no
S 420	Social Security "Lockbox"	no
S 27	Campaign Finance Reform	yes
S 27	Soft Money Cap	yes
H Con Res 83	Prescription Drug Benefit/Tax Cuts	no
H Con Res 83	Fiscal 2002 Budget Reconciliation	no
H Con Res 83	Funding for Environmental Programs	yes
H Con Res 83	"Marriage Penalty" Tax	no
H Con Res 83	Disabilities Education Act Funding	yes
S I	School Renovation and Construction	yes
HR 1836	Estate Tax	yes
HR 1836	Head Start	yes
HR 1836	Tax Cut Reconciliation Bill	no
S I	School Vouchers	no
S I	Boy Scouts/Anti-Discrimination	no
S 1052	Patients' Bill of Rights	yes
HR 2299	NAFTA/Mexican Trucks	yes
S 1438	Military Base Closures	yes
HR 2944	Fiscal 2002 District of Columbia Appropriations	yes

- (1) **Ashcroft Attorney General Confirmation.** Confirmation of President Bush's nomination of John Ashcroft of Missouri to serve as

U.S. Attorney General. Confirmed 58-42. Feb. 1, 2001. A no vote is a +.

- (2) **SJ Res 6. Ergonomics Rule Disapproval.** Passage of a joint resolution to reverse the ergonomics workplace safety rule submitted by the Clinton Administration's Labor Department. Passed 56-44. March 6, 2001. A no vote is a +.
- (3) **S 420. Social Security "Lockbox".** Domenici (R-NM) motion to waive the Budget Act in order to ensure that the Social Security surplus is used only to pay down the public debt until Social Security reform legislation is enacted. The bill would also ensure that the surplus in the Medicare Hospital Insurance Trust Fund is used only to pay down the public debt until Medicare reform legislation is enacted. Motion rejected 52-48 (a three-fifths majority vote - 60 - is required to waive the Budget Act.) March 13, 2001. A no vote is a +.
- (4) **S 27. Campaign Finance Reform.** McCain (R-AZ) motion to kill the Hatch (R-UT) amendment requiring unions and corporations to obtain permission from individual dues-paying workers or shareholders before spending money on political activities. The Hatch amendment was intended as a "poison pill" that, if passed and attached to the campaign finance reform bill, would destroy any chances the full reform bill had of passage. The Hatch amendment would also require corporations and unions to disclose information regarding the funds spent on political activities. Motion agreed to 69-31. March 21, 2001. A yes vote is a +.
- (5) **S 27. Soft Money Cap.** McCain (R-AZ) motion to kill a Hagel (R-NE) amendment to limit at \$60,000 per year soft money contributions by individuals, political action committees, corporations

and unions to national and state political party committees. The Hagel amendment would render the underlying reform bill's ban on soft money ineffective. Motion agreed to 60-40. March 27, 2001. A yes vote is a +.

- (6) **H Con Res 83. Prescription Drug Benefit/Tax Cuts.** Grassley (R-IA) amendment to reserve \$300 billion over 10 years to create a Medicare prescription drug benefit and overhaul the program. This amendment was a response to Democratic legislation which would have allocated \$311 billion for the benefit and not allowed the benefit's funding to come from the Medicare Hospital Trust Fund Surplus. Adopted 51-50, with Vice President Cheney casting a "yea" vote. April 3, 2001. A no vote is a +.
- (7) **H Con Res 83. Fiscal 2002 Budget Reconciliation.** Domenici (R-NM) amendment to instruct the Senate Finance Committee to report two reconciliation bills to the Senate that would reduce revenue levels by not more than the President's proposed \$1.6 trillion tax cut, and include a \$60 billion economic stimulus package for fiscal 2001. Adopted 51-49. April 5, 2001. A no vote is a +.
- (8) **H Con Res 83. Funding for Environmental Programs.** Corzine (D-NJ) amendment to increase funding for a wide variety of environmental programs by \$50 billion and set aside \$50 billion for debt reduction. The increases would be offset by reductions in the proposed tax cut. Rejected 46-54. April 5, 2001. A yes vote is a +.
- (9) **H Con Res 83. "Marriage Penalty" Tax.** Hutchison (R-TX) amendment to increase the proposed tax cut by \$69 billion for fiscal 2002-2011 in an effort to eliminate the co-called marriage penalty. Adopted 51-50, with Vice President Cheney casting a "yea" vote. April 5, 2001. A no vote is a +.

- (10) **H Con Res 83. Disabilities Education Act Funding.** Breaux (D-LA) amendment to redirect \$70 billion from the proposed tax cut to funding for the Individuals with Disabilities Education Act (IDEA) over 10 years. Adopted 54-46. April 5, 2001. A yes vote is a +.
- (11) **S I. School Renovation and Construction.** Harkin (D-IA) amendment to authorize \$1.6 billion for fiscal 2002 and such sums as necessary for each fiscal year between 2003 and 2006 for the construction and renovation of public elementary and secondary school buildings. Rejected 49-50. May 16, 2001. A yes vote is a +.
- (12) **HR 1836. Estate Tax.** Dorgan (D-ND) amendment to strike the estate-tax repeal provision and repeal the estate tax in 2003 for only all qualified family-owned farms and businesses. It also would reduce the top estate-tax rate bracket to 45 percent. Rejected 43-56. May 21, 2001. A yes vote is a +.
- (13) **HR 1836. Head Start.** Kennedy (D-MA) amendment to condition the reductions in the marginal income-tax rate on full funding for Head Start programs. Motion rejected 45-54. May 22, 2001. A yes vote is a +.
- (14) **HR 1836. Tax Cut Reconciliation Bill.** Adoption of the conference report on the bill to reduce taxes by \$1.35 trillion through fiscal 2011 through income tax rate cuts, relief of the "marriage penalty," phase-out of the federal estate tax, doubling of the child tax credit, and new incentives for retirement savings. A new 10 percent tax rate would be created retroactive to January 1. The bill would double the \$5000-per-child tax credit by 2010 and make it refundable, raise the estate tax exemption to \$1 million in 2002 and repeal the tax in 2010, increase the standard deduction for married couples to double

that of singles over five years, beginning in 2005, and increase annual contributions limits for Individual Retirement Accounts. The bill's provisions would expire December 31, 2010. Adopted 58-33. May 26, 2001. A no vote is a +.

- (15) **S I. School Vouchers.** Gregg (R-NH) amendment to create a demonstration program in 10 school districts to provide public school children with federal funds (vouchers) to transfer to another public school or a private school, including religious schools. The amendment would authorize \$50 million for fiscal 2002 and subsequent necessary sums for the next six fiscal years. Rejected 41-58. June 12, 2001. A no vote is a +.
- (16) **S I. Boy Scouts/Anti-Discrimination.** Helms (R-NC) amendment to withhold federal education funds from public elementary and secondary schools that bar the Boy Scouts of America from using school facilities. The targeted schools bar the Boy Scouts because the organization discriminates against gay men. Adopted 51-49. June 14, 2001. A no vote is a +.
- (17) **S 1052. Patients' Bill of Rights.** Passage of the bill to provide federal patient protections and allow patients to appeal a health maintenance organization's (HMO) decision on coverage and treatment. It also would allow patients to sue health insurers in state courts over quality-of-care claims and, at the federal level, over administrative or non-medical coverage disputes. Passed 59-36. June 29, 2001. A yes vote is a +.
- (18) **HR 2299. NAFTA/Mexican Trucks.** Shelby (R-AL) motion to uphold a border truck inspection program which allows Mexican trucks to receive three-month permits if they pass safety inspections. The motion also upholds a grant of \$60 billion to the Transportation

Department and various agencies. Motion agreed to 65-30. July 27, 2001. A yes vote is a +.

(19) **S 1438. Military Base Closures.** Warner (R-VA) motion to authorize an additional round of U.S. military base realignment and closures in 2003. Motion agreed to 53-47. September 25, 2001. A yes vote is a +.

(20) **HR 2944. Fiscal 2002 District of Columbia Appropriations.** Passage of the bill to provide \$408 million for the District of Columbia in fiscal 2002, including funds for the city's courts and corrections system and \$16.1 million for an emergency response plan following the September 11 attacks. The bill also would approve a \$7.2 billion budget for the District. Passed 75-24. November 7, 2001. A yes is a +.

REFERENCES

J. De Leeuw. Homogeneity Analysis of Pavings. URL <http://jackman.stanford.edu/ideal/MeasurementConference/abstracts/homPeig.pdf>. August 2003.

DEPARTMENT OF STATISTICS, UNIVERSITY OF CALIFORNIA, LOS ANGELES, CA 90095-1554

DEPARTMENT OF SOCIOLOGY, UNIVERSITY OF CALIFORNIA, LOS ANGELES, CA 90095-

E-mail address, Jan de Leeuw: deleeuw@stat.ucla.edu

URL, Jan de Leeuw: <http://gifi.stat.ucla.edu>

E-mail address, Paulette Lloyd: plloyd@ucla.edu

URL, Paulette Lloyd: http://www.soc.ucla.edu/grads.php?lid=1676&display_one=1