Scalarea şi analiza dimensională a datelor de tip ordine preferenţială: modele *unfolding*

Andrei Gheorghiţă*
Universitatea „Lucian Blaga" din Sibiu

Abstract:
This article explores the use of unidimensional unfolding for the scaling and dimensional analysis of preferential choice data. Although rather rarely used in the social sciences, preferential choice data allow the researchers to acquire a meaningful insight into the mechanisms of human choice, namely the identification of relevant criteria for individual choices. Unfolding models are discussed in the larger framework of Coombs' theory of data, with a focus on general purposes of scaling models (identification of underlying sources of variation, measurement, and data reduction). A simulation of unidimensional unfolding on empirical data is extensively considered.

Key words: unfolding, scaling models, dimensional analysis, preferential choice data

Scalare şi dimensionalitate

Termenul generic de *scalare* acoperă o plajă largă de modele şi metode de analiză statistică ce îşi propun identificarea surselor de variaţie în datele empirice cu care operează cercetătorul. Ele au în comun faptul că se bazează pe o metaforă geometrică (Weisberg, 1974: 743), altfel spus asumă posibilitatea unei modelări geometrice a informaţiilor provenind din observaţii empirice. Tehnicile de scalare sunt folosite în cercetarea empirică pentru cel puţin trei categorii de raţiuni (Coombs, 1964; Weisberg, 1974; McIver şi Carmines, 1981; Bădescu, 1999). În primul rând, ele pot fi folosite pentru a descrie structura unor date empirice, prin identificarea dimensiunilor latente care stau în spatele variaţiei acestora. Scopul este practic unul explorator, iar rezultatul poate fi identificarea unei dimensiuni latente (unidimensionalitate) sau a mai multora (multidimensionalitate) care stau la baza variaţiei. Spre exemplu, putem găsi o dimensiune ideologică (stânga-dreapta) care să explice variaţiile în intenţiile de vot ale indivizilor pentru o mulţime dată de partide politice sau putem identifica simultan şi alte dimensiuni latente (cum ar fi libertarianism-autoritar). În al doilea rând, scalarea poate avea drept obiectiv măsurarea comportamentelor/atitudinilor individuale, prin dezvoltarea unei scale pe care indivizilor li se conferă scoruri. Aceasta este situaţia unui sociolog care construieşte o scală a statusului socio-economic pe care să poziţioneze indivizii pornind de la anumite variabile atitudinale şi comportamentale (McIver şi Carmines, 1981: 9). În al treilea rând, finalitatea scalării poate fi în mod explicit reducerea complexităţii datelor, prin diminuarea numărului de variabile în condiţiile păstrării aceleiaşi structuri a datelor (Bădescu, 1999: 254).

Modelele de scalare pot fi folosite pentru a scala persoane (indivizii primesc scoruri pe o scală prin combinarea răspunsurilor/valorilor individuale), stimuli (caracteristici ale obiectelor primesc scoruri prin scalarea valorilor percepute de

* Adresa de corespondenţă: andrei.gheorghita@ulbsibiu.ro.

către mai mulţi indivizi) sau răspunsuri ale indivizilor la stimuli (răspunsul va varia în funcţie de stimul, dar şi de poziţia proprie a individului în legătură cu caracteristica în cauză). Pentru fiecare situaţie, doar anumite modele de scalare sunt potrivite.

Totodată, aşa cum observam anterior, modelele de scalare pot genera soluţii unidimensionale sau multidimensionale, în funcţie nu doar de natura datelor, ci şi de natura metodei/tehnicii în discuţie. Vorbim despre scalare unidimensională atunci când este întemeiată presupunerea unei unice dimensiuni fundamentale latente în spatele unui set de date empirice (McIver şi Carmines, 1981: 13). Atunci când soluţia unidimensională pare a fi insuficientă pentru a explica structura datelor, apare necesitatea unei abordări multidimensionale (Weisberg, 1974: 745), implicând plasarea obiectelor într-un spaţiu cu mai multe dimensiuni, definite pe baza proprietăţilor acestora (McIver şi Carmines, 1981: 13). Astfel, *dimensionalitatea datelor* trebuie definită ca „numărul de surse distincte şi interesante de variaţie între obiecte” (Jacoby, 1991: 27). Analiza dimensionalităţii implică din partea cercetătorului considerarea simultană a două categorii de aspecte: (a) numărul caracteristicilor după care diferă între ele obiectele mulţimii şi (b) relavanţa acestora pentru obiectivele investigaţiei ştiinţifice.

Tipuri de date

Sugeram anterior că există o legătură între tipul de date aflate la dispoziţia cercetătorului şi tehnicile de scalare care li se pot aplica. Pentru o dezvoltare a subiectului, o scurtă incursiune în *teoria datelor* este absolut necesară. O contribuţie esenţială la dezvoltarea acestui domeniu îi revine lui Clyde H. Coombs, care publică în 1964 extrem de influentul studiu „A Theory of Data”, bazându-se exclusiv pe o interpretare geometrică a datelor. Conform acestei viziuni, orice observaţie empirică presupune o comparaţie între două entităţi, care pot fi reprezentate ca puncte într-un spaţiu. Poziţionarea reciprocă a punctelor în acel spaţiu depinde de modul în care analistul alege să interpreteze natura comparaţiei dintre respectivele entităţi (Jacoby, 1991: 15). Spre exemplu, observaţia potrivit căreia „candidatul A este social-democrat” ar putea fi transpusă geometric printr-o poziţionare apropiată în spaţiu a punctelor corespunzătoare „candidatului A” şi statutului de „social-democrat”. Sau, observaţia potrivit căreia „candidatul A obţine mai multe voturi decât candidatul B” poate implica o reprezentare geometrică în care punctul „A” ocupă în spaţiu o poziţie mai apropiată de una din extremităţi decât punctul „B”. Astfel, se poate observa că natura datelor determină modul în care modelul geometric va fi construit pornind de la observaţiile empirice (Jacoby, 1991: 15). În acest mod, reprezentarea geometrică va funcţiona ca strategie de modelare abstractă a datelor.

În viziunea lui Coombs, datele pot fi clasificate pornind de la două criterii:

> „(1) Cele două elemente constitutive ale perechii pot fi extrase din *mulţimi diferite*[1] (spre exemplu, un student şi o întrebare la un test, un consumator şi un produs, un stimul şi un răspuns) sau pot fi extrase din *aceeaşi mulţime* (spre exemplu, studentul A şi studentul B care completează acelaşi test, marca X şi marca Y dintr-un set de produse de consum, stimulul 1 şi stimulul 2 dintr-un pachet experimental).

[1] Este vorba despre entităţi de naturi diferite.

> (2) Comparaţia dintre entităţile din pereche implică o *relaţie de dominanţă* (spre exemplu, un student răspunde corect la întrebarea unui test, un cobai de laborator parcurge un labirint mai repede decât un altul, magnitudinea unui stimul depăşeşte un anumit nivel marcat pe un instrument de măsură) sau o *relaţie de proximitate* (spre exemplu, două produse conţin un acelaşi ingredient, un membru al unui grup îl selectează pe altul ca partener de activitate, un răspuns specific însoţeşte un stimul dat)." (Coombs, Dawes şi Tversky, 1970, apud. Jacoby, 1991: 16).

Cele două distincţii dihotomice pot fi transpuse cu uşurinţă în plan grafic[2]. Intersectând cele două criterii, Coombs defineşte patru mari tipuri de date: (a) *date de tip stimul unic*; (b) *date de tip comparaţie de stimuli*; (c) *date de similaritate* şi (d) *date de tip alegere preferenţială* (Figura 1).

În categoria *datelor de tip stimul unic* intră, în general, informaţiile provenind din categoria măsurătorilor. Elementele sunt de acelaşi tip, iar relaţia dintre ele una de dominaţie. Cele două mulţimi de puncte vor reprezenta obiectele măsurate şi, respectiv, unităţile de înregistrare ale instrumentului de măsură. Spre exemplu, dacă un obiect are 10 centimetri, atunci lungimea sa va domina toate semnele unei rigle situate mai jos de cel care indică 10 cm şi va fi dominată de toate semnele superioare acestuia. Datele de acest tip pot fi modelate apelând la scale sumative, scale cumulative, analiză factorială şi analiză cluster (Bădescu, 1999: 255).

		Perechea de puncte din observaţie	
		Aceeaşi mulţime	Mulţimi diferite
Relaţia dintre punctele din pereche	Dominaţie	*Date de tip comparaţie de stimuli*	*Date de tip stimul unic*
	Proximitate	*Similarităţi*	*Alegere preferenţială*

Figura 1: *Cele patru tipuri de date conform teoriei lui Coombs*
Sursă: Jacoby, 1991:18.

Datele de tip comparaţie de stimuli sunt specifice situaţiilor de comparare a unor obiecte similare pe baza unei caracteristici comune. Astfel, individului îi este prezentată o mulţime de stimuli, cerându-i-se să îl selecteze pe acela care are mai mult/mai puţin din respectiva caracteristică decât ceilalţi (McIver şi Carmines, 1981: 12). Se instituie astfel o relaţie de dominaţie după acea caracteristică, ce poate fi modelată geometric ca ordonare a punctelor corespunzând celor doi sau mai mulţi stimuli după o axă numerică (Jacoby, 1991: 19). În această categorie de date intră

[2] Dacă elementele perechii provin din seturi diferite, spaţiul se va numi *reunit* ('*joint space*'), deoarece va conţine două seturi diferite de puncte (câte unul pentru fiecare entitate „observată" repetat). Dacă elementele provin din acelaşi set, atunci punctele se vor găsi într-un *spaţiu obiect* ('*object space*'). Dacă entităţile sunt legate printr-o relaţie de dominanţă, aceasta se va reflecta în ordonarea punctelor în spaţiu. Punctul asociat entităţii dominante va fi amplasat într-o poziţie mai apropiată de o extremitate (de obicei, mai mare numeric) după o dimensiune. Proximitatea este modelată ca distanţă între puncte. Cu cât cele două obiecte sunt mai apropiate, cu atât distanţa dintre punctele asociate lor devine mai mică şi vice-versa. (Jacoby, 1991: 17)

situaţiile în care individului i se cere să aleagă cea mai dulce dintre două tablete de ciocolată ori cel mai economic dintre două autoturisme, fără a fi implicată în vreun fel în alegere preferinţa sa. Altfel spus, selecţia individuală va reflecta exclusiv diferenţele dintre stimuli.

În cazul *datelor de similaritate*, avem de-a face cu situaţia în care perechi de stimuli de acelaşi tip sunt evaluaţi din perspectiva a cât de similari sunt. Modelarea se face numai sub aspectul distanţei şi nicidecum al ordinii. Cu cât stimulii sunt mai similari, cu atât apropierea (proximitatea) dintre ei creşte şi invers. Astfel, dacă asemănarea dintre autoturismele A şi B este mai mare decât cea dintre autoturismele A şi C, aceasta înseamnă că A seamănă mai mult cu B decât cu C, deci punctul A (ca reprezentare geometrică) va fi situat mai aproape de B decât de C. În procesul de evaluare a similarităţilor, indivizilor nu li se oferă nici un fel de indicaţii asupra dimensiunilor sau criteriilor pe care să le folosească (McIver şi Carmines, 1981: 12), identificarea lor ulterioară fiind la latitudinea cercetătorului. Pentru datele de acest tip se aplică modelele de scalare multidimensională.

A patra categorie de date, în viziunea lui Coombs, o reprezintă cele *de tip alegere preferenţială*. Se aplică comportamentelor de tip preferenţial: cu cât unui individ îi „place" mai mult sau „preferă" un anumit stimul, cu atât este mai mare apropierea (proximitatea) dintre subiect şi acel stimul. Din punct de vedere geometric, proximităţile sunt reprezentate ca distanţe între puncte într-un spaţiu reunit. Creşterea proximităţii dintre un subiect şi un stimul corespunde unei descreşteri a distanţei dintre punctul asociat subiectului şi punctul asociat stimulului. Informaţia conţinută într-o singură alegere preferenţială nu oferă nici un fel de informaţie în legătură cu ordonarea subiectului şi a stimulilor în spaţiu (Jacoby, 1991: 18-22). O situaţie de acest gen ar apărea, spre exemplu, atunci când unui individ X i s-ar cere să aşeze în ordinea preferinţelor candidaţii A, B, C şi D, care concurează pentru funcţia de primar al unei localităţi. Cu cât unul dintre candidaţi ar fi plasat mai sus în ordinea preferinţelor, cu atât punctul asociat acestuia va fi mai apropiat în spaţiul reunit de punctul corespunzător subiectului, însă identificarea criteriului/criteriilor de ordonare reprezintă o provocare adresată cercetătorului. Pentru scalarea acestui tip de date se aplică cel mai frecvent modelele numite *unfolding*.

Principii generale ale modelelor unfolding

Metoda *unfolding* încearcă să gestioneze datele de tip *alegere preferenţială*. Este vorba despre situaţii în care unor subiecţi umani li se oferă un set de stimuli şi aceştia trebuie să-i ordoneze în funcţie de preferinţe. Prin *unfolding*, cercetătorul încearcă să identifice factorii care îi determină pe subiecţi în alegerea ordinii de preferinţe. Altfel spus, ce au oamenii în minte atunci când fac o astfel de alegere? Poate exista *o dimensiune* care gestionează alegerea (*unfolding* unidimensional) sau *mai multe* (*unfolding* multidimensional).

Există două strategii de a culege date de tip alegere preferenţială. Dacă avem un număr redus de stimuli, subiecţilor li se poate cere să îi aşeze într-o ordine de preferinţe. Dacă însă avem un număr mai mare de stimuli, aceştia vor fi prezentaţi subiecţilor în perechi, cerându-li-se să îşi exprime preferinţa pentru un stimul din fiecare pereche posibilă. Pentru respondent, exprimarea preferinţelor în perechi este, în mod evident, mai facilă. Pentru cercetător însă, există un risc semnificativ ca o bună parte din subiecţi să nu respecte logica tranzitivităţii şi, astfel, o bună parte a datelor culese să fie distorsionate. Dintr-un atare motiv, este probabil ca, operând cu acelaşi set de stimuli, datele obţinute să fie diferite, într-o mai mare sau mai mică măsură.

În analiza dimensională a datelor prin tehnica unfoldind, fiecare stimul va fi reprezentat ca un punct pe o axă de tipul *îmi place – nu îmi place* (‘*like – dislike*’). Fiecare axă de acest gen va reprezenta, în analiză, câte o dimensiune a alegerii, altfel spus ca o sursă de variabilitate a preferinţelor. De partea cealaltă, fiecare subiect va fi reprezentat pe acea axă printr-un *punct ideal*, reprezentând punctul de maximă preferinţă al individului în raport cu dimensiunea respectivă. Semnificaţia punctului ideal pentru un individ este următoarea – este punctul în care dacă ar fi poziţionat un stimul, atunci acel stimul ar fi preferat de respectivul individ în raport cu oricare alt stimul. El „măsoară” poziţionarea fiecărui individ în raport cu acea dimensiune.

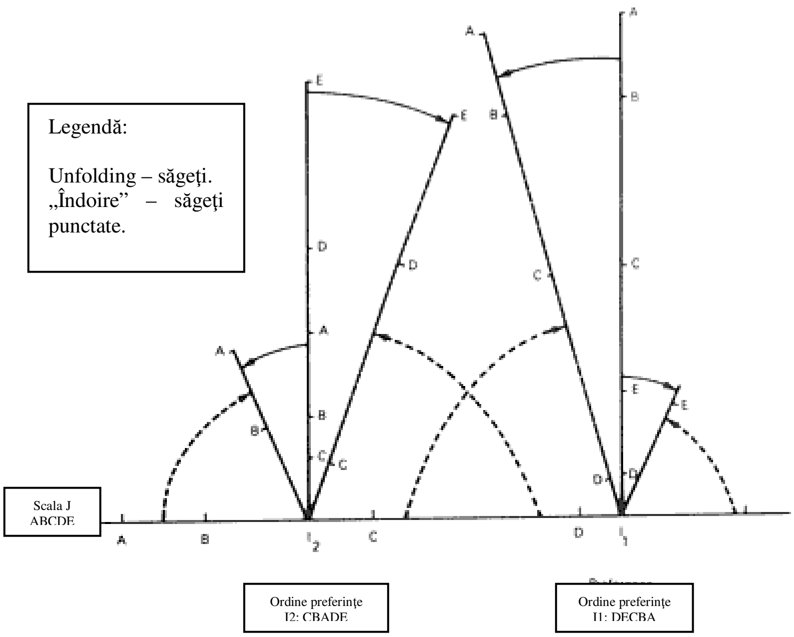


Figura 2: *Cum pot fi „dezdoite” două scale de preferinţe individuale (I1 şi I2) pentru a forma scala comună (J), respectiv cum poate fi „îndoită” scala comună (J) pentru a reprezenta scalele preferinţelor individuale (I1 şi I2)*
Notă: A, B, C, D, E reprezintă stimulii individuali, iar I1 şi I2 poziţiile a doi subiecţi.
Sursă: McIver şi Carmines, 1981: 73.

Modelul unfolding urmează o regulă simplă: dacă un subiect preferă stimulul A stimulului B, atunci distanţa de la punctul său ideal la A va fi mai mică decât distanţa de la punctul său ideal la B:

$$\text{dist}(I, A) < \text{dist}(I, B)$$

Denumirea de „unfolding” (‘*to fold*’ din limba engleză poate fi tradus prin *a îndoi* şi ‘*to unfold*’ prin *a dezdoi*) asociată modelului sugerează tocmai principiul de modelare geometrică a datelor de tip alegere preferenţială. Astfel, asumpţia fundamentală este aceea potrivit căreia *n* axe ale preferinţelor individuale (în care punctul ideal al individului este plasat la una dintre extremităţi) ar putea fi „dezdoite” pe o unică scală a preferinţelor (soluţia unidimensională)[3], cu care toate axele să fie consistente şi care să reflecte dimensiunea folosită pentru ordonarea stimulilor (Figura 2). Pornind de la regula enunţată mai sus, vor face obiectul reprezentării grafice pe scala unică toate punctele asociate celor *k* (stimuli), respectiv *n* (indivizi).

Simularea unui caz de *unfolding* unidimensional

Pentru a exemplifica tehnica de unfolding[4], să imaginăm situaţia în care unui număr de *n* subiecţi (reprezentaţi pe linii în matricea datelor) li se cere să aşeze cinci localităţi în ordinea preferinţei de stanilire a domiciliului (deci cinci stimuli reprezentaţi pe coloane în matricea datelor): Iaşi (I), Bacău (B), Suceava (S), Oneşti (O) şi Târgu Ocna (T). Intrările în celulele matricei vor reprezenta poziţia stimului de pe coloană în ordinea de preferinţe a individului de pe linie: 1 reprezintă prima preferinţă, 2 a doua, 3 a treia ş.a.m.d. Rezultă o matrice de proximitate: cu cât este mai mic un scor, cu atât va fi mai „apropiată” individului acea localitate (Figura 3).

Fie cazul unidimensional, în care asumăm premisa că ordinile preferenţiale au în background o unică dimensiune evaluativă. Altfel spus, fiecare subiect compară stimulii folosind acelaşi criteriu. Evident, vom opera cu o singură axă de tipul *îmi place – nu îmi place*, reprezentând unica dimensiune latentă. Într-o primă etapă, vom reprezenta pe axă punctele asociate celor cinci stimuli (I, B, S, O şi T). La acest nivel intervine o primă problemă – în ce ordine vom reprezenta stimulii pe axa *îmi place – nu îmi place* ? Pentru aceasta, primul pas îl constituie analiza ordinilor de preferinţe exprimate de către subiecţi.

	I	**B**	**S**	**O**	**T**	**Ordine**
Individ 1	3	1	2	4	5	BSIOT
Individ 2	5	4	2	1	3	OSTBI
Individ 3	5	3	2	1	4	OSBTI
Individ 4	1	2	3	4	5	IBSOT
Individ 5	4	1	2	3	5	BSOIT
Individ *n*	...	...	...	...	...	.....

Figura 3: *Exemplu de matrice a datelor pentru unfolding*

Pentru *k* stimuli vor exista întotdeauna *k!* ordonări posibile. Dacă însă modelul nostru asumat unidimensional este corect (implicit respectă şi regula tranzitivităţii), în practică vor fi posibile mai puţine ordonări, adică doar:

$$\frac{k(k-1)}{2}+1$$

[3] Sau pe câteva, pentru soluţia multidimensională.

[4] Simularea urmează succesiunea etapelor de modelare *unfolding* prezentate de William J. Jacoby în cursul „Scaling and Dimensional Analysis”, Ann Arbor, MI: ICPSR Program in Social Statistics, 2002. Materialele de curs sunt disponibile online la adresa: http://polisci.msu.edu/jacoby/icpsr/scaling.

ordini preferenţiale. În cazul nostru, pentru 5 stimuli, ar trebui să existe cel mult 11 ordini preferenţiale valide.

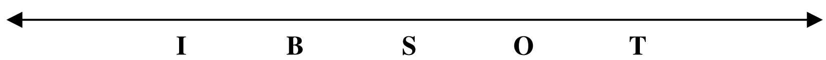
În această etapă, dintre ordinile preferenţiale exprimate de către subiecţi, urmează să identificăm stimulii care apar cel mai adesea pe ultima poziţie, deci care sunt cel mai frecvent cel mai puţin preferaţi. Să exemplificăm pe cazul ideal, în care operăm cu 11 ordini preferenţiale consistente cu un model unidimensional, aşa cum sunt ele exprimate de 11 subiecţi:

BSIOT
OSTBI
OSBTI
IBSOT
BSOIT
TOSBI
BISOT
SBOTI
SBOIT
SOBTI
OTSBI

Pentru un model unidimensional pur, pe ultima poziţie nu se vor putea afla decât 2 stimuli, în cazul nostru T şi I. Dacă în ordinile preferenţiale exprimate de către subiecţi apar mai mult de 2 stimuli pe ultimele poziţii exprimate individual, înseamnă fie că modelul este multidimensional, fie că se încalcă regula tranzitivităţii. Odată identificaţi cei 2 stimuli cel mai puţin preferaţi, am identificat *cele două extreme ale axei* – pentru cazul nostru T şi I.

În a doua etapă se caută ordinile preferenţiale care au cei doi stimuli pe primul loc. Odată identificate, acestea ar trebui să fie reciproc imagini în oglindă. Studiind răspunsurile celor 11 respondenţi, ne oprim asupra următoarelor ordini preferenţiale: IBSOT şi TOSBI. Dacă, pentru un model aparent unidimensional, identificăm două sau mai multe perechi de ordini preferenţiale imagini în oglindă având stimulii extremi pe prima poziţie, atunci este foarte probabil ca modelul nostru să fie totuşi multidimensional.

Revenind la exemplul nostru, observăm că IBSOT şi TOSBI sunt imagini în oglindă. Acestea vor fi ordinile de poziţionare a stimulilor pe axă, pornind de la o extremă a sa ori de la cealaltă. Reprezentând grafic axa, obţinem:



Trebuie făcută o observaţie importantă: punctul ideal al oricărui respondent care a ales ordinea preferenţială IBSOT va fi situat pe axă la stânga lui I, în timp ce punctul ideal al oricărui respondent care alege TOSBI va fi situat pe axă la dreapta lui T.

În a treia etapă, pentru a uşura operaţiile ulterioare, simbolurile asociate stimulilor, în ordinea poziţionării lor pe axă, sunt înlocuite cu litere ordonate alfabetic. Astfel, reetichetăm stimulii după cum urmează:

I (Iaşi) devine	A
B (Bacău) devine	B
S (Suceava) devine	C
O (Oneşti) devine	D

T (Târgu Ocna) devine E.

Pentru etapele următoare, vom opera cu un concept nou, şi anume acela de *punct median*. Punctul median este acel punct care împarte un interval în două subintervale de dimesiuni identice. Spre exemplu, pe axa stimulilor, intervalul dintre punctele A şi B va avea punctul median notat AB. Pentru cinci stimuli, obţinem combinaţii de 5 luate câte 2 puncte mediane, deci 10 puncte mediane: AB, AC, AD, AE, BC, BD, BE, CD, CE şi DE, care vor împărţi axa în 11 *subintervale*. Importanţa acestor subintervale este enormă, prin faptul că unui punct ideal plasat oriunde pe un astfel de subinterval îi corespunde *o ordine preferenţială şi numai una*.

Următoarea etapă constă în identificarea ordinii preferenţiale asociate fiecărui subinterval. La acest nivel intervine un subterfugiu metodologic, bazat tocmai pe caracteristica ordinală (la nivel de alfabet) a etichetelor A, B, C, D şi E. În acest mod, pentru fiecare *ordine preferenţială* putem afla câte puncte mediane lasă la stânga sa şi câte la dreapta sa. Cum procedăm? Fie prima ordine preferenţială, BSIOT, devenită după recodare BCADE. Începem cu prima literă din ordinea preferenţială şi ne deplasăm spre dreapta numărând literele care o precedă în alfabet. În cazul lui B o precedă doar A, deci îi asociem un scor de 1:

B 1

Continuăm cu a doua literă din ordinea preferenţială şi repetăm procedura. În cazul lui C, mergând spre dreapta, observăm că o precedă doar A, deci îi asociem un scor de 1:

C 1

Repetăm pentru a treia literă, A, şi apoi pentru a patra, D:

A 0
D 0

Totalizând aceste scoruri, obţinem un scor de 2 asociat ordinii preferenţiale, ce trebuie interpretat ca *numărul de puncte mediane situate la stânga*.

Procedura se reia pentru toate ordinile preferenţiale consistente cu modelul unidimensional:

BCADE=2
DCEBA=8
DCBEA=7
ABCDE=0
BCDAE=3
EDCBA=10
BACDE=1
CBDEA=5
CBDAE=4
CDBEA=6
DECBA=9

Aşezând crescător ordinile preferenţiale în funcţie de numărul de puncte mediane la stânga, putem să le ordonăm/poziţionăm pe axa stimulilor. De la stânga la dreapta, ordinile preferenţiale vor fi:

ABCDE	0
BACDE	1
BCADE	2
BCDAE	3
CBDAE	4

CBDEA	5
CDBEA	6
DCBEA	7
DCEBA	8
DECBA	9
EDCBA	10

Rămâne să identificăm punctele mediane ce despart subintervalele asociate ordinilor preferenţiale de mai sus. Aceasta se va realiza urmărind perechile de litere inversate în fiecare două ordini preferenţiale consecutive. Astfel, pentru primele două ordini preferenţiale, ABCDE şi BACDE, observăm inversarea perechii de litere AB (respectiv BA). Aceasta înseamnă că subintervalele definite prin ordinile de preferinţe ABCDE şi, respectiv, BACDE vor fi separate de punctul median AB. Procedura se repetă pentru toate celelalte subintervale:

Ordini de preferinţe	**Numărul de puncte mediane la stânga**	**Puncte mediane intersectate (de graniţă)**
ABCDE	0	
		AB
BACDE	1	
		AC
BCADE	2	
		AD
BCDAE	3	
		BC
CBDAE	4	
		AE
CBDEA	5	
		BD
CDBEA	6	
		CD
DCBEA	7	
		BE
DCEBA	8	
		CE
DECBA	9	
		DE
EDCBA	10	

Figura 4: *Ordini de preferinţe şi puncte mediane*

În aceste condiţii, putem ordona pe axă punctele mediane. Ordonând aceste puncte, introducem informaţie spaţială în ordinea stimulilor, în sensul că le estimăm nu doar ordinea, ci şi „distanţele" care îi separă după dimensiunea analizată. Această etapă precede momentul culminant al analizei de tip *unfolding*, când are loc extragerea de *informaţie metrică* din date de tip alegere preferenţială.

Informaţia metrică va fi conţinută de punctele mediane „îmbrăţişate". Porţiunea liniară de axă cuprinsă între două puncte poate fi interpretată ca un segment. Acesta este definit printr-o anumită lungime. Între punctele mediane putem considera existenţa unor segmente pe axa dimensională. Atunci când un astfel de segment este conţinut într-un altul mai lung, de aceeaşi natură, cele patru puncte mediane care definesc extremităţile celor două segmente sunt *puncte mediane îmbrăţişate* (‘*enveloping midpoints*’). Pentru a uşura înţelegerea conceptului, poate fi avută în vedere analogia cu poezia şi utilizarea rimei îmbrăţişate.

Înainte de a continua demonstraţia, vom introduce o convenţie de notare. Dacă AB desemnează punctul median al intervalului cuprins între A şi B, $\overline{AB}$ va desemna segmentul de linie delimitat prin punctele A şi B. În plus, imaginând axa ca fiind „imprimată” cu o unitate de măsură pentru lungime, fiecărui punct îi corespunde o informaţie metrică de poziţionare (abscisă).

Primul pas constă în listarea exhaustivă a segmentelor ce pot fi definite pe axă de către punctele asociate celor cinci stimuli. Vom avea:

a) segmente de dreaptă primare: $\overline{AB}$, $\overline{BC}$, $\overline{CD}$ şi $\overline{DE}$;
b) segmente de dreaptă secundare (unesc două segmente de dreaptă primare): $\overline{AC}$, $\overline{BD}$ şi $\overline{CE}$;
c) segmente de dreaptă terţiare (unesc trei segmente de dreaptă primare): $\overline{AD}$ şi $\overline{CE}$.

Al doilea pas constă în compararea lungimilor segmentelor de dreaptă. În acest scop vom recurge la compararea punctelor mediane şi nu a segmentelor de dreaptă în sine. Vom putea compara la acest moment doar segmentele de dreaptă care îndeplinesc două condiţii, adică: nu au o extremitate comună şi nu se suprapun. În continuare, vom parcurge toate combinaţiile de două segmente posibile:

1) $\overline{AB}$ vs. $\overline{BC}$
Comparaţia nu este posibilă, deoarece segmentele au o extremitate comună.

2) $\overline{AB}$ vs. $\overline{CD}$
Comparăm punctele mediane definite de literele interioare şi exterioare din perechea de segmente.[5] Urmărim ordinea punctelor mediane obţinută anterior:

AD < BC

Operând cu abscise, obţinem:

AD=$\frac{A+D}{2}$ şi BC=$\frac{B+C}{2}$

Deci:

$$\frac{A+D}{2} < \frac{B+C}{2}$$

A+D < B+C

[5] Pentru fiecare segment de dreaptă literele trebuie ordonate alfabetic.

Scadem la stânga şi la dreapta un C:
A+D-C < B+C-C
A+D-C < B

Scadem la stânga şi la dreapta un A:
A-A+D-C < B-A
D-C < B-A

D şi C sunt extremităţile segmentului $\overline{CD}$, deci diferenţa celor două abscise va da lungimea segmentului. La fel pentru segmentul $\overline{AB}$. De unde rezultă:
$\overline{CD} < \overline{AB}$

3) $\overline{AB}$ vs. $\overline{DE}$
AE < BD
$$\frac{A+E}{2} < \frac{B+D}{2}$$
A+E < B+D
E < B+D-A
E-D < B-A
$\overline{DE} < \overline{AB}$

4) $\overline{BC}$ vs. $\overline{CD}$
Comparaţia nu este posibilă, deoarece segmentele au o extremitate comună.

5) $\overline{BC}$ vs. $\overline{DE}$
CD < BE
C+D < B+E
C+D-B < E
C-B < E-D
$\overline{BC} < \overline{DE}$

6) $\overline{CD}$ vs. $\overline{DE}$
Comparaţia nu este posibilă, deoarece segmentele au o extremitate comună.

7) $\overline{AC}$ vs. $\overline{BD}$
Comparaţia nu este posibilă, deoarece segmentele se suprapun parţial.

8) $\overline{BD}$ vs. $\overline{CE}$
Comparaţia nu este posibilă, deoarece segmentele se suprapun parţial.

9) $\overline{AC}$ vs. $\overline{CE}$
Comparaţia nu este posibilă, deoarece segmentele au o extremitate comună.

10) $\overline{AD}$ vs. $\overline{BE}$
Comparaţia nu este posibilă, deoarece segmentele se suprapun parţial.

Segmentele de dreaptă terţiare au o extremitate comună sau se suprapun parţial cu orice alt segment. Din acest motiv le vom exclude din toate comparaţiile ulterioare.

11) $\overline{AB}$ vs. $\overline{AC}$
Comparaţia nu este posibilă, deoarece segmentele se suprapun parţial.

12) $\overline{AB}$ vs. $\overline{BD}$
Comparaţia nu este posibilă, deoarece segmentele au o extremitate comună.

13) $\overline{AB}$ vs. $\overline{CE}$
BC < AE
B+C < A+E
B-A < E-C
$\overline{AB} < \overline{CE}$

14) $\overline{BC}$ vs. $\overline{AC}$
Comparaţia nu este posibilă, deoarece segmentele se suprapun parţial.

15) $\overline{BC}$ vs. $\overline{BD}$
Comparaţia nu este posibilă, deoarece segmentele se suprapun parţial.

16) $\overline{BC}$ vs. $\overline{CE}$
Comparaţia nu este posibilă, deoarece segmentele au o extremitate comună.

17) $\overline{CD}$ vs. $\overline{AC}$
Comparaţia nu este posibilă, deoarece segmentele au o extremitate comună.

18) $\overline{CD}$ vs. $\overline{BD}$
Comparaţia nu este posibilă, deoarece segmentele se suprapun parţial.

19) $\overline{CD}$ vs. $\overline{CE}$
Comparaţia nu este posibilă, deoarece segmentele se suprapun parţial.

20) $\overline{DE}$ vs. $\overline{AC}$
AE < CD
A+E < C+D
A+E-D < C
E-D < C-A
$\overline{DE} < \overline{AC}$

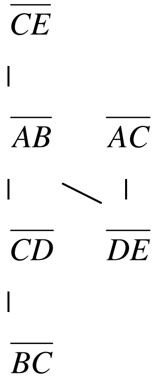
21) $\overline{DE}$ vs. $\overline{BD}$
Comparaţia nu este posibilă, deoarece segmentele au o extremitate comună.

22) $\overline{DE}$ vs. $\overline{CE}$
Comparaţia nu este posibilă, deoarece segmentele se suprapun parţial.

Recapitulând, am obţinut:

$$\overline{CD} < \overline{AB}$$
$$\overline{DE} < \overline{AB}$$
$$\overline{BC} < \overline{DE}$$
$$\overline{AB} < \overline{CE}$$
$$\overline{DE} < \overline{AC}$$

Aşezăm segmentele într-o diagramă, în care un segment de dreaptă este mai lung cu cât este mai sus amplasat.



În plus faţă de informaţia transmisă prin poziţionare, în diagramă, relaţiei de comparaţie între perechile de segmente este reprezentată printr-o linie care uneşte respectivele segmente.

Cum obţinem informaţia metrică? Răspunsul este unul simplu: pe diagramă ataşăm ca lungimi pentru segmentele reprezentate orice seturi de numere care sunt consistente cu modelul. A formula în termeni de „orice seturi de numere" poate părea o asumpţie cu o raţionalitate destul de neclară. Însă, poate fi demonstrat statistic că, orice seturi de numere consistente cu modelul am ataşa, fiecare set va fi foarte aproape de o *transformare liniară*[6] a oricărui alt set de numere consistent cu modelul (un coeficient de corelaţie r >0,9).

Să pornim de la cel mai scurt dintre segmentele comparate, $\overline{BC}$, şi să îi acordăm, în mod absolut arbitrar, valoarea 1:

$$\overline{BC} = 1$$

La un prim nivel, segmentul $\overline{BC}$ este mai scurt decât $\overline{DE}$ şi $\overline{CD}$, segmente care, însă, nu pot fi comparate între ele. Deci, singura condiţie impusă de model

---

6 De tipul $y_i = a + bx_i$.

(diagramă) este ca cele două segmente să fie mai lungi decât $\overline{BC}$. Să asumăm, din nou arbitrar, că ele iau valorile:

$$\overline{CD}=4$$

$$\overline{DE}=5$$

În aceste condiţii, putem calcula şi lungimea segmentului $\overline{CE}$, care va fi suma lungimilor segmentelor enumerate anterior:

$$\overline{CE}=\overline{CD}+\overline{DE}=4+5=9$$

Deci:

$$\overline{CE}=9$$

În acest moment, lipsesc evaluări metrice pentru două segmente, $\overline{AB}$ şi $\overline{AC}$. Din punct de vedere al lungimii, asupra segmentului $\overline{AB}$ acţionează două categorii de constrângeri din partea modelului: (a) trebuie să fie mai lung decât $\overline{CD}$ şi $\overline{DE}$ (deci mai lung de 5 unităţi), dar mai scurt decât suma lor, $\overline{CE}$ (deci mai scurt de 9 unităţi) şi (b) trebuie să fie mai scurt decât $\overline{AC}$, deoarece este doar o parte a acestuia ($\overline{AC}=\overline{AB}+\overline{BC}$).

Răspunzând acestor constrângeri, să îi acordăm segmentului $\overline{AB}$ lungimea de 6 unităţi:

$$\overline{AB}=6$$

În aceste condiţii:

$$\overline{AC}=\overline{AB}+\overline{BC}=6+1=7$$

Deci:

$$\overline{AC}=7$$

Recapitulând, am asumat:

$\overline{CE}$ (9)

$\overline{AC}$ (7)

$\overline{AB}$ (6)

$\overline{CD}$ (4) $\overline{DE}$ (5)

$\overline{BC}$ (1)

Aceste date ne permit să localizăm metric cei cinci stimuli (A, B, C, D şi E) pe axa reprezentând dimensiunea studiată. Deoarece A este stimulul aflat cel mai la stânga[7], pentru a ne uşura calculele, putem să îl asumăm ca origine, altfel spus:

$$A=0$$

Punctul B va fi la 6 unităţi de A, deoarece $\overline{AB}=6$, iar C la 7 unităţi, deoarece $\overline{AC}=7$. Deci:

$$B=6$$

[7] Sau cel mai la dreapta, dacă am porni de la imaginea în oglindă.

C=7

Punctul D va fi situat la 4 unităţi spre dreapta de punctul C, deoarece $\overline{CD}$=4, iar E la 9 unităţi, deoarece $\overline{CE}$=9. De unde:

D=C+4=7+4=11 D=11

şi:

E=C+9=7+9=16 E=16

Reprezentând grafic, obţinem:

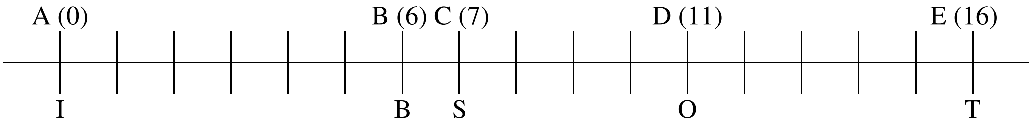


Figura 5: *Poziţionarea stimulilor pe scală în baza informaţiei metrice*

Analizând dimensionalitatea pentru cazul alegerilor preferenţiale, ne propunem să identificăm criteriile la care recurg subiecţii atunci când evaluează respectivii stimuli. În spatele fiecărei dimensiuni stă *un criteriu* sau un *cumul de criterii*, care va fi/vor fi cu atât mai semnificativ/-e cu cât ordinile preferenţiale consistente cu acea dimensiune sunt mai frecvente.

Revenind la modelul nostru, presupus unidimensional, următorul pas constă în identificarea dimensiunii care „gestionează” ordinile preferenţiale analizate. Ca şi în cazul altor tehnici de scalare şi analiză dimensională, criteriile rezultate din analiză nu se dezvăluie singure, nu sunt evidente şi, nicidecum, transparente. Este nivelul la care intervin maximal experienţa, flerul şi imaginaţia cercetătorului. De asemenea, trebuie subliniat că nu întotdeauna putem avea certitudinea că interpretarea aleasă este optimă şi, nicidecum, unică.

O privire atentă asupra rezultatelor prezentei analize va indica faptul că, cel mai probabil, criteriul utilizat de către respondenţii noştri în evaluarea celor cinci categorii de oraşe este cel al *mărimii localităţii* (cu toate implicaţiile conexe asupra stilului de viaţă, resurselor şi oprtunităţilor), mergând de la un oraş foarte populat cum este Iaşi (I) şi până la dimensiunile extrem de restrânse ale unui oraş predominant turistic cum este Târgu Ocna (T). Mai mult, informaţia metrică extrasă ne permite să comparăm *diferenţele percepute* de respondenţi între cele cinci localităţi. Spre exemplu, sub aspectul implicaţiilor mărimii oraşului, diferenţa percepută între Bacău (B) şi Suceava (S) este de patru ori mai mică decât cea dintre Suceava (S) şi Oneşti (O). Se cuvine subliniat încă o dată că relativa simplitate a interpretării criteriului de ordonare este dată de caracterul particular al datelor „fabricate” şi totodată ideale pe care s-a construit acest exemplu. În practică însă, datele reale nu vor furniza niciodată certitudinea relevanţei şi unicităţii soluţiei găsite.

Ultima etapă în demersul nostru rezidă în localizarea (metrică) a punctelor ideale ale respondenţilor după dimensiunea analizată. Astfel, valoarea numerică asociată fiecărei ordini preferenţiale alese va fi media aritmetică a valorilor punctelor mediane care mărginesc subintervalul asociat respectivei ordini preferenţiale. Să exemplificăm cu situaţia unui individ care alege ordinea preferenţială BACDE. Subintervalul asociat acestei ordini preferenţiale este delimitat de punctele mediane AB şi AC, localizate numeric la 3 şi, respectiv, 3,5 unităţi de origine (A):

$$AB=\frac{A+B}{2}=\frac{0+6}{2}=3$$

şi:

$$AC=\frac{A+C}{2}=\frac{0+7}{2}=3,5$$

Potrivit regulii indicate anterior, punctul ideal al unui individ care alege ordinea preferenţială BACDE va fi situat la:

$$I_{BACDE}=\frac{AB+AC}{2}=\frac{3+3,5}{2}=3,25$$

unităţi de origine. Aceste valori numerice ale puctelor ideale individuale, odată identificate, pot face obiectul altor analize de ordin cantitativ. Dacă replicăm procedura pentru ceilalţi 10 indivizi, obţinem informaţia metrică privind poziţionarea pe axă a tuturor punctelor ideale corespunzătoare celor 11 ordini preferenţiale (Tabel 1).

Ordine de preferinţe	Puncte mediane învecinate		Punct ideal	Ordine stimuli corespunzătoare
BCADE	AC=3,5	AD=5,5	4,50	BSIOT
DCEBA	BE=11	CE=11	11,25	OSTBI
DCBEA	CD=9	BE=11	10	OSBTI
ABCDE	A=0	AB=3	1,50	IBSOT
BCDAE	AD=5,5	BC=6,5	6	BSOIT
EDCBA	DE=13,5	E=16	14,75	TOSBI
BACDE	AB=3	AC=3,5	3,25	BISOT
CBDEA	AE=8	BD=8,5	8,25	SBOTI
CBDAE	BC=6,5	AE=8	7,25	SBOIT
CDBEA	BD=8,5	CD=9	8,75	SOBTI
DECBA	CE=11,5	DE=13,5	12,5	OTSBI

Tabel 1: *Informaţie metrică privind punctele ideale corespunzătoare celor 11 ordini preferenţiale consistente cu modelul unidimensional*

Odată extrasă informaţia metrică, se poate obţine reprezentarea geometrică fidelă a punctelor asociate stimulilor, punctelor mediane şi, respectiv, intervalelor de ordine preferenţială pe axa identificată, în cazul nostru cea a *mărimii localităţii* (Figura 6), practic ultima etapă a modelării unfolding.

Sintetizând, modelele de unfolding unidimensional îşi propun scalarea simultană a stimulilor şi indivizilor, astfel încât ordinile preferenţiale ale diferiţilor indivizi să fie consistente cu poziţiile alocate respectivilor indivizi pe un continuum (McIver şi Carmines, 1981: 84-85). Această metodă are la bază principiul proximităţii în modelarea geometrică, materializat în reflectarea proximităţii psihologice între stimul şi individ (relevată de ordonarea preferinţelor) în apropiere între punctele corespunzătoare acestora într-un spaţiu unidimensional. În ciuda utilizării mai degrabă rare a acestei metode de scalare în spaţiul ştiinţelor sociale, există o serie de avantaje ale implementării sale, ce pot fi cu greu ignorate. Avem în vedere, în primul rând, capacitatea de a integra stimuli şi poziţii relative ale unor indivizi într-un continuum unidimensional înzestrat cu consistenţă internă, facilitând astfel identificarea criteriilor latente implicate în alegeri individuale. În al doilea rând, modelele unfolding oferă o cale consistentă de extragere a informaţiei metrice din date de tip ordinal, prin „măsurarea” poziţiilor individuale şi ale stimulilor de-a lungul dimensiunii latente identificate. Odată extrasă această informaţie metrică, ea poate fi integrată în analize suplimentare de ordin statistic.

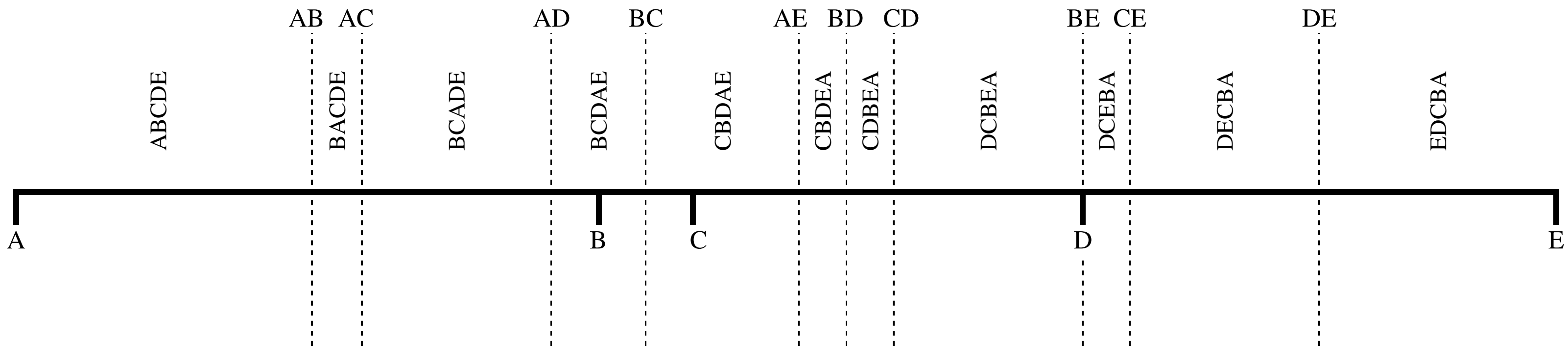


Figura 6: *Unfolding unidimensional: puncte asociate stimulilor, puncte mediane şi intervale de ordine preferenţială*

Bibliografie:

Bădescu, Gabriel (1999) „Metode de reducere a datelor”, în Traian Rotariu (coord.); Gabriel Bădescu; Irina Culic; Elemér Mezei şi Cornelia Mureşan, *Metode statistice aplicate în ştiinţele sociale*, Iaşi: Polirom, pp. 254-271.

Coombs, Clyde H. (1964) *A Theory of Data*, New York: John Wiley and Sons.

Culic, Irina (2004) *Metode avansate în cercetarea socială. Analiza multivariată de interdependenţă*, Iaşi: Polirom.

Jacoby, William J. (1991) *Data Theory and Dimensional Analysis*, Newbury Park: Sage Publications.

Jacoby, William J. (1982) „Unfolding the Party Identification Scale: Improving the Measurement of an Important Concept”, în *Political Methodology*, vol. 8, no. 2, pp. 33-59.

McIver, John P. şi Edward G. Carmines (1981) *Unidimensional Scaling*, Newbury Park: Sage Publications.

Norpoth, Helmut (1979) „The Parties Come to Order! Dimensions of Preferential Choice in the West German Electorate, 1961-1976”, în *The American Political Science Review*, vol. 73, no. 3, pp. 724-736.

Weisberg, Herbert F. (1974) „Dimensionland: An Excursion into Spaces”, în *American Journal of Political Science*, vol. 18, no. 4, pp. 743-776.