

Un indicator informațional pentru variabilele categoriale ale unei anchete. Notă de cercetare

Gabriel Sticlaru*

Institutul de Cercetare a Calității Vieții, Academia Română

Abstract: The statistical procedures used for data processing have to be considerate depending on measurement variables level: nominal, ordinal, interval or ratio type. Here are presented some examples for ordinal variables, where calculated mean is misleading and gives a wrong interpretation. For the categorical variables, we propose an indicator resulted from the informational entropy, as a measure of the responses distribution in a survey. Calculations are made for a Likert scale, survey data from 1991 to 2010. The indicator allows us to make a global characterization of the distributions, items ranking, dynamic comparisons and patterns identification.

Keywords: *Scales of measurement; Informational Entropy and Energy; Likert Scale.*

Cuvinte-cheie: *scale de măsurare; entropie și energie informațională; Scală Likert.*

Introducere

Analiza datelor măsurate pe o scală ordinală nu este întotdeauna transparentă. În multe cazuri, sunt tratate ca scale de interval, pentru a beneficia de avantajul unor tehnici statistice complexe (bazate pe medie, coeficienți de corelație și împrăștiere). Probabil că în multe cazuri scala nu satisface ipotezele unei scale de interval, de exemplu cerința ca intervalele dintre treptele scalei să fie egale.

De regulă, pentru date ordinale, se pot folosi procedurile nonparametrice bazate pe frecvențe, ranguri, mediană sau amplitudine, dar nu și procedurile parametrice bazate pe medie și deviație standard.

Tratarea datelor ordinale ca date măsurate pe o scală de interval (sau raport), fără examinarea prealabilă a valorilor și obiectivelor cercetării, poate conduce la concluzii greșite.

Sunt multe cazuri în care interpretarea datelor pe baza frecvențelor răspunsurilor este distorsionată, de exemplu, prin prezentarea mediilor itemilor scalei.

Vom exemplifica cu rezultatele obținute din ancheta „Diagnoza Calității Vieții 2010”.

Vom considera o scală Likert cu 5 variante de răspuns:

1. Foarte scăzută
2. Scăzută
3. Satisfăcătoare
4. Ridicată
5. Foarte ridicată

Itemii scalei sunt (conform notației din chestionar):

D35 Accesibilitatea formelor de învățământ dorite de dvs. și de cei apropiați;

D36 Posibilitatea de a obține un loc de muncă pe măsura intereselor și a capacității (pentru dvs. și cei apropiați);

* Institutul de Cercetare a Calității Vieții, Academia Română, Calea 13 Septembrie nr. 13, București.
E-mail: gabriel.sticlaru@gmail.com.

D37 Securitatea dvs. de acasă, de pe stradă etc.

În tabelul de mai jos sunt prezentate frecvențele răspunsurilor (în procente) și media fiecărui item al scalei.

Table 1: Exemplu în care media itemilor nu este relevantă

Itemi	Variante de răspuns					Media
	1	2	3	4	5	
D35	5,20	23,50	41,37	26,71	3,21	2,99
D36	39,52	37,10	17,34	5,69	0,35	1,90
D37	4,06	18,06	42,44	31,72	3,72	3,13

Dacă 41,4% dintre respondenți percep D35 ca „satisfăcător”, există două minorități semnificative (23,5% și 26,7%) cu opinii contrare („scăzut”, respectiv „ridicat”). O analiză pe baza mediei de 3,0 ar explica pentru D35, că în medie, opinia a fost „satisfăcător”.

Pentru D36, 37,1% au selectat varianta „scăzut”, dar două grupuri importante (39,5%, respectiv 17,3%), caracterizează drept „foarte scăzut”, respectiv „satisfăcător”. Valoarea mediei de 1,90 ar indica opinia „scăzut”.

Pentru D37, opinia a 42,4% dintre respondenți este „satisfăcător”, dar observăm și două grupuri importante (18,1% și 31,7%) cu opinii contrare („scăzut”, respectiv „ridicat”). O interpretare cu media de 3,1 arată un acord general față de varianta „satisfăcător”.

Există și exemple cu răspunsurile concentrate la capetele scalei (1, respectiv 5) și cu media la mijloc (varianta 3), cu o eronată explicație față de distribuția frecvențelor.

Metodologie

În cele ce urmează ne interesează să obținem o ierarhie a itemilor unei scale, bazată numai pe frecvențele răspunsurilor.

Vom considera o construcție agregată care este măsurată pe o scală cu n trepte (de obicei 3, 5, 7 sau 9), agregarea fiind cu m

itemi. Scala poate fi nominală, ordinală sau de interval.

Pentru a măsura diversitatea sau concentrarea în cadrul unui sistem, putem folosi entropia sau energia informațională, concepte introduse de Claude Shannon, respectiv Octav Onicescu.

Pe baza conceptului de entropie din teoria informației, vom construi un indicator care va permite caracterizarea distribuției frecvențelor răspunsurilor. Am preferat Entropia informațională, deoarece, spre deosebire de Energia informațională, este un indicator mediu, ponderat.

Pentru fiecare item, considerăm proporția răspunsurilor (frecvențe relative) corespunzătoare celor n variante: (p_1, \dots, p_n) , cu $\sum_{i=1}^n p_i = 1$.

Pentru această distribuție, entropia informațională este indicatorul ponderat:

$$H = - \sum_{i=1}^n p_i \log p_i$$

Dacă luăm logaritmi în baza n , atunci entropia maximă va fi $H_{\max} = \log_n n = 1$.

Entropia fiind simetrică, vom cere ca itemii scalei să fie corelați pozitiv. Distribuțiile (70%, 20%, 10%) și (10%, 20%, 70%) au aceeași entropie, dar prezintă o corelație negativă. Această cerință este în general satisfăcută în cazul unui construct care folosește un set de itemi orientați în același sens pe scală.

Pe baza entropiei informaționale, vom defini un indice relativ, pe care-l vom numi *gradul de diferențiere (de concentrare)*:

$$D = H_{\max} - H = 1 - H.$$

Distribuția cu proporțiile răspunsurilor egale este considerată uniformă, nediferențiată, cu $D = 0$.

Pe măsură ce răspunsurile se diferențiază (concentrează), indicele D crește (se apropie de 1). În tabele vom prezenta procentual atât frecvențele relative, cât și indicele de diferențiere (D).

Rezultate și discuție

Vom folosi datele culese prin anchetele privind „Diagnoza Calității Vieții” pentru anii 1991–1999, 2003, 2006 și 2010. Vom analiza scala introdusă prin întrebarea:

În perioada următoare, de ce anume vă temeți cel mai mult, pentru dvs. și familia dvs.?

Tabel 2: Scală Likert cu 3 trepte și 5 itemi

Item		Mult	Puțin	Deloc
1.	Șomaj	1	2	3
2.	Conflicte sociale	1	2	3
3.	Lipsa de securitate personală datorită criminalității	1	2	3
4.	Creșterea prețurilor	1	2	3
5.	Impozite	1	2	3

Scala va fi aplicată pentru perioada 1991–1999, 2003, 2006, 2010.

Tabel 3: Date pentru anul 2010

Item	%				Rang			
	1	2	3	D	T1	T2	T3	D
1	42,0	23,1	34,9	2,6	3	3	1	4
2	26,6	43,6	29,8	2,1	5	2	2	5
3	28,0	45,2	26,8	2,7	4	1	3	3
4	90,0	8,2	1,8	66,1	1	5	5	1
5	87,5	10,3	2,2	60,6	2	4	4	2

În coloanele T1, T2, T3 am trecut ierarhizarea celor 5 itemi conform treptei de intensitate 1, 2, 3. De exemplu, conform scalei T1, respondenții se tem cel mai mult de (4) Creșterea prețurilor, apoi de (5) Impozite, (1) Șomaj, (3) Insecuritate personală și (2) Conflicte sociale, deci în ordinea (4, 5, 1, 3, 2).

După scala T2 itemii se ierarhizează în ordinea (3, 2, 1, 5, 4), iar după T3 în ordinea (1, 2, 3, 5, 4).

Dacă ierarhiile T1, T2 și T3 sunt particulare, în funcție de valoarea pe scală, clasa-mentul itemilor după entropia informațională (indicele de diferențiere D) ia în calcul toate cele 3 valori ale scalei, deci este o **ierarhie medie**, care ține cont de proporțiile răspunsurilor în toate cele 3 trepte. Astfel,

putem considera că din punct de vedere al temerii, ordinea dată de scala informațională este (4, 5, 3, 1, 2).

Putem vizualiza abaterile clasificărilor graduale T1, T2, T3 față de scala informațională.

Tabel 4: Date pentru anul 2006

Item	%				Rang			
	1	2	3	D	T1	T2	T3	D
1	25,0	22,6	52,4	7,1	3	2	2	4
2	10,6	36,5	52,9	14,2	4	1	1	3
3	10,6	36,5	52,9	14,2	4	1	1	3
4	84,5	11,5	4,0	52,7	1	4	4	1
5	80,5	13,8	5,7	44,4	2	3	3	2

Tabel 5: Date pentru anul 2003

Item	%				Rang			
	1	2	3	D	T1	T2	T3	D
1	43,4	17,9	38,7	5,6	3	3	2	3
2	18,5	41,6	39,9	5,0	5	1	1	4
3	26,0	41,6	32,4	1,7	4	2	3	5
4	87,9	9,5	2,6	60,9	1	5	5	1
5	82,5	13,2	4,3	48,7	2	4	4	2

Tabel 6: Date pentru anul 1999

Item	%				Rang			
	1	2	3	D	T1	T2	T3	D
1	45,2	23,5	31,3	3,3	3	3	1	4
2	34,0	41,1	24,9	1,8	5	1	2	5
3	40,0	39,7	20,3	3,8	4	2	3	3
4	91,3	6,3	2,4	68,2	1	5	5	1
5	87,4	9,3	3,3	58,8	2	4	4	2

Tabel 7: Date pentru anul 1998

Item	%				Rang			
	1	2	3	D	T1	T2	T3	D
1	44,7	21,3	34,0	3,9	3	2	1	5
2	21,2	51,1	27,7	6,4	5	1	2	4
3	36,1	51,1	12,8	11,4	4	1	3	3
4	78,7	17,0	4,3	43,2	1	3	5	1
5	74,5	17,0	8,5	33,5	2	3	4	2

Tabel 8: Date pentru anul 1997

Item	%				Rang			
	1	2	3	D	T1	T2	T3	D
1	41,0	21,7	37,3	3,0	3	4	2	4
2	16,0	40,8	43,2	7,0	5	1	1	3
3	32,3	40,4	27,3	1,2	4	2	3	5
4	89,3	9,5	1,2	65,6	1	5	5	1

Item	%				Rang			
	1	2	3	D	T1	T2	T3	D
5	69,8	22,2	8,0	28,3	2	3	4	2

Tabel 9: Date pentru anul 1996

Item	%				Rang			
	1	2	3	D	T1	T2	T3	D
1	35,8	23,4	40,8	2,3	3	3	1	4
2	19,4	40,1	40,5	4,3	5	2	2	3
3	30,1	41,2	28,7	1,2	4	1	3	5
4	89,8	8,1	2,1	65,1	1	5	5	1
5	72,4	19,1	8,5	30,9	2	4	4	2

Tabel 10: Date pentru anul 1995

Item	%				Rang			
	1	2	3	D	T1	T2	T3	D
1	31,1	21,7	47,2	4,5	3	3	1	3
2	20,1	35,9	44,0	4,3	5	2	2	4
3	25,0	39,6	35,4	1,6	4	1	3	5
4	87,7	8,2	4,1	59,0	1	5	5	1
5	80,0	14,1	5,9	43,2	2	4	4	2

Tabel 11: Date pentru anul 1994

Item	%				Rang			
	1	2	3	D	T1	T2	T3	D
1	42,5	23,4	34,1	2,6	3	3	2	4
2	22,3	38,4	39,3	2,7	5	2	1	4
3	35,2	42,4	22,4	2,9	4	1	3	3
4	89,0	8,7	2,3	63,3	1	5	5	1
5	80,3	11,8	7,9	42,8	2	4	4	2

Tabel 15: Ierarhizare după indicele de diferențiere

Item	Rang											
	1991	1992	1993	1994	1995	1996	1997	1998	1999	2003	2006	2010
1	5	3	3	4	3	4	4	5	4	3	4	4
2	4	4	4	4	4	3	3	4	5	4	3	5
3	3	5	5	3	5	5	5	3	3	5	3	3
4	1	1	1	1	1	1	1	1	1	1	1	1
5	2	2	2	2	2	2	2	2	2	2	2	2

Se observă că pentru întreaga perioadă, temerea cea mai mare este acordată itemilor (4) Creșterea prețurilor și (5) Impozite. Oamenii se tem în continuare de (3) Insecuritate personală, (1) Somaj și (2) Conflicte sociale.

Tabel 16: Valorile indicatorului de diferențiere

Item	Anul											
	1991	1992	1993	1994	1995	1996	1997	1998	1999	2003	2006	2010
1	1,0	5,3	9,4	2,6	4,5	2,3	3,0	3,9	3,3	5,6	7,1	2,6
2	5,0	1,1	3,7	2,7	4,3	4,3	7,0	6,4	1,8	5,0	14,2	2,1

Tabel 12: Date pentru anul 1993

Item	%				Rang			
	1	2	3	D	T1	T2	T3	D
1	29,4	17,0	53,6	9,4	4	3	1	3
2	20,4	39,4	40,2	3,7	5	1	2	4
3	36,0	35,9	28,1	0,6	3	2	3	5
4	89,8	6,4	3,8	63,8	1	5	5	1
5	76,4	15,0	8,6	36,2	2	4	4	2

Tabel 13: Date pentru anul 1992

Item	%				Rang			
	1	2	3	D	T1	T2	T3	D
1	34,9	19,0	46,1	5,3	3	3	1	3
2	27,1	39,7	33,2	1,1	5	1	2	4
3	34,4	36,1	29,5	0,3	4	2	3	5
4	86,7	9,7	3,6	57,3	1	5	5	1
5	75,3	17,4	7,3	35,6	2	4	4	2

Tabel 14: Date pentru anul 1991

Item	%				Rang			
	1	2	3	D	T1	T2	T3	D
1	38,0	26,7	35,3	1,0	5	3	1	5
2	39,8	41,7	18,5	5,0	4	1	2	4
3	50,9	33,0	16,1	8,6	3	2	3	3
4	82,4	14,2	3,4	49,7	1	5	5	1
5	76,9	17,5	5,6	39,2	2	4	4	2

Spre deosebire de ranguri, indicele de diferențiere permite identificarea unor paternuri pentru perioada considerată.

Item	Anul											
	1991	1992	1993	1994	1995	1996	1997	1998	1999	2003	2006	2010
3	8,6	0,3	0,6	2,9	1,6	1,2	1,2	11,4	3,8	1,7	14,2	2,7
4	49,7	57,3	63,8	63,3	59,0	65,1	65,6	43,2	68,2	60,9	52,7	66,1
5	39,2	35,6	36,2	42,8	43,2	30,9	28,3	33,5	58,8	48,7	44,4	60,6

În perioada 1991–1997, temerea față de (4) – Creșterea prețurilor este din ce în ce mai mare, în raport cu (5) – Impozite. Mine-riadele din 1990–1991 și represiunea mișcărilor sociale din 1990 determină pentru anul 1991 o temere crescută pentru (3) – Insecuritate personală și (2) – Conflicte sociale. Se observă că în preajma alegerilor, este o temere crescută față de Șomaj (1). Paternul se modifică în anul 2010 spre o tipologie cu două grupuri, cu intensități foarte diferite de apreciere. Itemii (4) și (5) sunt pe același plan de apreciere a temerii (foarte ridicată), iar itemii (1), (2) și (3) pe un plan secundar (foarte scăzută).

Concluzii

Interpretările pe baza mediei aritmetice în cazul variabilelor măsurate pe o scală ordinală trebuie făcute cu precauție. Încrederea acordată mediei aritmetice sporește dacă, de exemplu, vom înlocui scala Likert cu 5 trepte cu o scală cu 7 sau 9 trepte. Dacă vom folosi însă o scală de intensitate, de tip „termometru”, nivelul de măsurare poate fi considerat ca fiind de interval.

Pentru exemplificare, vom considera scala de intensitate introdusă prin întrebarea:

D74. În orice societate, unii oameni se consideră bogați, alții se consideră săraci.

Având în vedere numerotarea de la 1 la 10, dvs. unde vă situați?

1	2	3	4	5	6	7	8	9	10
Sărac									Bogat

Interpretări interesante se pot obține dacă comparăm mediile pe categoriile unei alte variabile, de exemplu D1 – Sex sau D6 – Pregătire școlară.

În Tabelul 17 de mai jos se observă că media răspunsurilor nu diferă în funcție de sex.

Distribuția mediilor (de la 2,35 la 5,65) este corelată pozitiv cu nivelul de studii.

Cei fără școală se consideră săraci, iar cei cu studii superioare se consideră suficient de bogați.

Tabel 17: Media în 2010 pentru D74 în funcție de D1

D1 – Sexul	Media
1. Masculin	4,40
2. Feminin	4,50
Total eșantion	4,46

Tabel 18: Media în 2010 pentru D74 în funcție de D6

D6 – Pregătirea școlară	Media
1. Fără școală	2,35
2. Școală generală neterminată	3,67
3. Școală generală absolvită	3,78
4. Școală profesională (școală de arte și meserii)	4,38
5. Liceu	4,69
6. Școală postliceală	5,08
7. Studii superioare	5,65
Total eșantion	4,46

Indicatorul informațional introdus de noi este bazat pe frecvențele relative și poate fi aplicat variabilelor categoricale ale anchetei, cu număr fix de categorii. Din cauza simetriei indicatorilor informaționali (entropie sau energie informațională), poziția în cadrul distribuției nu este preluată în calcule, astfel că distribuția (10%, 15%, 75%) are același indicator informațional pentru orice permutare.

Putem face însă comparații în dinamică pentru aceeași variabilă și observa tendin-

țele, în cazul în care nu apar răsturnări spectaculoase de procente între două momente de timp succesive, cum ar fi (10%, 20%, 70%) și (70%, 10%, 20%).

De asemenea, putem ierarhiza itemii unui construct, măsurat pe o scală Likert, la care în general itemii scalei sunt corelați pozitiv (corelația rangurilor).

Toate calculele noastre statistice au folosit datele primare din cercetarea „Diagnoza Calității Vieții” care s-a realizat anual în perioada 1990–1999 și în anii 2003, 2006,

2010, fiind finanțată prin programul de cercetare al Institutului de Cercetare a Calității Vieții.

Tabelele prezentate exprimă în dinamică o evaluare globală a scalei, la nivelul fiecărui item. Calculele indicatorului pentru perioada 1991–2010 au permis caracterizarea distribuției răspunsurilor, ierarhizări, comparații și identificarea unor paternuri în opiniile respondenților.

Bibliografie

Onicescu, O. (1966) Energie informationelle. *Comptes Rendues l'Academie des Sciences*, Paris, A, 263.

Shannon, C.E. (1948) A mathematical theory of communication part I and II. *Bell Syst. Tech. Journal*, 27, 379–423, 623–656.