

## Statistică socială si SPSS

- ghid pentru curs practic -

Profesor: Lect.dr. Ioan Hosu

Asistent si tutore: asist.drd. Mihai Deac

Contact: [mhdeac@yahoo.com](mailto:mhdeac@yahoo.com), program de consultatii – luni, 14-16, vineri, 10-11

La nivelul cercetării de orice tip, inclusiv cea socio-umană, elementele de statistică au devenit indispensabile. Revistele științifice internaționale publică în majoritatea situațiilor date empirice, care să susțină eventualele construcții teoretice, iar standardele acestora în ceea ce privește prelucrarea datelor cantitative sunt foarte ridicate. Astfel că, orice autor sau cercetător serios are nevoie de cunoștințe avansate de prelucrare a datelor empirice cantitative. Similar, în toate domeniile profesionale asociate științelor comunicării, creația și inspirația nu sunt acceptate fără argumente clare de ordin statistic. Departamentul de creație nu convinge pe nimeni dacă ideile pe care le prezintă nu se pliază pe unele date culese inițial de departamentul de cercetare.

Asta nu înseamnă că metodele calitative trebuie brusc neglijate. Este exagerat astăzi să te plasezi într-o abordare științifică pur cantitativă, dominată de pozitivism și de matematicizarea tuturor proceselor sociale. Nu toate științele se pot transforma în matematică, chiar dacă elementele cantitative dau un plus de forță oricărei argumentații. Ajungem la ideea că statistica este necesară, dar nu suficientă atunci când facem cercetare în domeniul socio-uman.

## Variabilele

Pentru că ne interesează să măsurăm aproape tot ce se poate măsura, caracteristicile ale căror trăsături vrem să le aflăm trebuie privite ca *variabile*. Acestea sunt de fapt categorii care urmează să fie măsurate. Variabilele sunt de cel puțin 4 tipuri:

- Nominale – sunt acele caracteristici sub formă de nume sau simboluri, care nu pot fi ordonate între ele. De exemplu: genul unei persoane este variabilă nominală cu valorile 1=masculin, 2=feminin. Însă aici valoarea 2 nu este neapărat superioară valorii 1. Un alt exemplu de variabilă nominală este *canalul de informație folosit cel mai des*. Valorile pot fi: 1-televiziunea, 2-internetul, 3-presa scrisă, 4-radio, 5-altele. Din nou valoarea 4 nu este superioară, nici inferioară unei alte valori din sir. Ordinea între valori poate fi schimbată oricând.
- Ordinale – sunt acele caracteristici care presupun o ordine naturală (intrinsecă) a valorilor. Educația este un exemplu de variabilă ordinală. Valorile sunt în ordine, de la 1-scoală primară, 2-gimnaziu, 3-scoală profesională, 4-liceu, 5-postliceală, 6-facultate etc. Totuși, pentru variabilele ordinale nu putem face împărțiri, deoarece distanțele

dintre trepte nu sunt egale (între scoală primară și gimnaziu nu este aceeași distanță ca între liceu și postliceală).

- De intervale – variabile ordinale, dar cu diferență egală între valori. De menționat că pentru aceste variabile nu există valoarea 0 absolută pe scală. De exemplu, coeficientul de inteligență este o variabilă de intervale.
- De rapoarte – variabile pur numerice, cum ar fi greutatea, vârsta, înălțimea.

Deseori, în practică, nu se face o diferențiere între variabile de intervale și de rapoarte, ambele fiind considerate variabile cantitative, numerale sau scalare.

### **Ipoteza**

O propoziție care stabilește o relație între două variabile și sensul acesteia. Acestea trebuie să apară în formă explicită în planul lucrării, în introducerea ei, dar și în prezentarea ei finală. Trebuie să fie cuantificabile, demonstrabile, să nu fie excesiv de generale și să nu prezinte niște adevăruri evidente (care nu mai au nevoie de o demonstrație). Munca în SPSS își propune de fapt testarea de ipoteze.

### **Introducerea datelor în SPSS**

Programul SPSS este împărțit în două ferestre: Data View și Variable View.

În **Variable View** se definesc variabilele cu care urmează să lucrăm. Astfel, fiecare întrebare dintr-un chestionar este introdusă ca variabilă. I se pune un nume generic, apoi i se alege tipul: de regulă *numeric* sau *string*. *String* se utilizează atunci când valorile nu pot fi exprimate în cifre, adică în cazul întrebărilor deschise. Se poate seta lărgimea coloanei, precum și numărul de zecimale. La *Label* se trece explicarea variabilei, numele ei detaliat. *Label*-ul apare ulterior pe toate tabelele și graficele care vor implica variabila respectivă. Cel mai important aspect în Variable View este introducerea fiecărei valori, cu codul care îi corespunde. De obicei, introducem fiecare variantă de răspuns a întrebării în ordine, de sus în jos, începând de la 1. În cazul unui non-răspuns de obicei introducem valoarea 99. La opțiunea *Missing* trebuie să trecem non-răspunsurile, adică valoarea 99. Aceste valori nu vor fi luate în considerare în calculul procentelor valide.

ATENȚIE – dacă nu selectăm valorile *Missing*, valoarea de tip 99 se va lua în calcul, iar o medie calculată pe acea variabilă va fi afectată de eroare. De exemplu, am în esantion 70 de persoane cu vârsta medie de 33 de ani și am un număr de 35 de persoane în același esantion, care nu și-au declarat vârsta, drept pentru care am marcat varianta 99. Dacă nu semnalăm 99 ca valoare lipsă, media de vârstă calculată de SPSS va fi în jur de 45 de ani, ceea ce desigur nu este corect.

La *Measure* trebuie să trecem tipul variabilei. Există 3 opțiuni: *Nominal*, *Ordinal* și *Scale*. În funcție de ce tip de variabilă este, vom ști ce operații putem face cu ea. Problema apare la variabilele nominale, pe care nu putem calcula medii, corelații ș.a.

The screenshot shows the SPSS Data Editor interface. A table lists variables with their names, types, widths, decimals, labels, values, missing values, columns, alignment, and measure. A 'Value Labels' dialog box is open, showing a list of values and their corresponding labels. The dialog box has fields for 'Value' and 'Value Label', and buttons for 'Add', 'Change', 'Remove', 'OK', 'Cancel', and 'Help'.

Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
1 plimban	Numeric	8	2	plimbări	(1.00, zilnic)	None	8	Right	Scale
2 mers_pej	Numeric	8	2	mers pe jos	(1.00, zilnic)	None	8	Right	Scale
3 sport_sa	Numeric	8	2					Right	Scale
4 sport_li	Numeric	8	2					Right	Scale
5 sport_in	Numeric	8	2					Right	Scale
6 sport_ec	Numeric	8	2					Right	Scale
7 sala_fit	Numeric	8	2					Right	Scale
8 sport_tv	Numeric	8	2					Right	Scale
9 sport_a	Numeric	8	2					Right	Scale
10 performa	Numeric	8	2					Right	Scale
11 sport1	String	88	0					Left	Nominal
12 varsta1	Numeric	8	2					Right	Scale
13 timp1	Numeric	8	2	perioada (nr de	None	None	8	Right	Scale
14 sport2	String	88	0	sport 2	None	None	8	Left	Nominal
15 varsta2	Numeric	8	2	vârsta la care	None	None	8	Right	Scale
16 timp2	Numeric	8	2	perioada (nr de	None	None	8	Right	Scale
17 sport_b	Numeric	8	2	am văzut sport	(1.00, acord to	None	8	Right	Scale
18 prieteni	Numeric	8	2	aveam prieteni	(1.00, acord to	None	8	Right	Scale
19 parinti	Numeric	8	2	parinții au venit	(1.00, acord to	None	8	Right	Scale
20 familie	Numeric	8	2	in familia mea	(1.00, acord to	None	8	Right	Scale
21 idoli	Numeric	8	2	aveam idoli pri	(1.00, acord to	None	8	Right	Scale
22 famili_a	Numeric	8	2	in familia mea	(1.00, acord to	None	8	Right	Scale
23 motiv	Numeric	8	2	cel mai import	(1.00, pentru a	None	8	Right	Scale
24 partener	Numeric	8	2	care sunt parte	(1.00, colegii d	None	8	Right	Scale
25 locatii	Numeric	8	2	in ce locatii pr	(1.00, pe teren	None	8	Right	Scale
26 locati_a	String	88	0	care alte locati	None	None	8	Left	Nominal
27 cheltuie	Numeric	8	0	cheltuieli, echi	None	None	8	Right	Nominal
28 acces	Numeric	8	0	probleme acce	(1, in foarte ma	None	8	Right	Nominal
29 oameni_n	Numeric	8	2	cunoastere oa	(1.00, foarte de	None	8	Right	Scale
30 sanatate	Numeric	8	2	probleme de s	(1.00, foarte de	None	8	Right	Scale
31 obosit	Numeric	8	2	obosit si fara c	(1.00, acord to	None	8	Right	Scale
32 stare_bi	Numeric	8	2	am o stare gen	(1.00, acord to	None	8	Right	Scale
33 increder	Numeric	8	2	am incredere i	(1.00, acord to	None	8	Right	Scale
34 capabil	Numeric	8	2	nu ma mai sim	(1.00, acord to	None	8	Right	Scale
35 stres	Numeric	8	2	ma simt eliber	(1.00, acord to	None	8	Right	Scale
36 frustrat	Numeric	8	2	sunt frustrat	(1.00, acord to	None	8	Right	Scale
37 tonus	Numeric	8	2	imi creste tonu	(1.00, acord to	None	8	Right	Scale
38 singur	Numeric	8	2	in viata trebuie	None	None	8	Right	Scale
39 solidari	Numeric	8	2	daca cei din ju	None	None	8	Right	Scale
40 reguli	Numeric	8	2	imi place sa re	None	None	8	Right	Scale
41 sarcina	Numeric	8	2	daca ma angaj	None	None	8	Right	Scale
42 efort	Numeric	8	2	sunt separat c	None	None	8	Right	Scale
43 echipa	Numeric	8	2	sunt un om car	None	None	8	Right	Scale
44 reguli2	Numeric	8	2	imi place sa in	None	None	8	Right	Scale
45 parere	Numeric	8	2	parerea celorla	None	None	8	Right	Scale
46 priete_a	Numeric	8	2	am multi priete	None	None	8	Right	Scale

În **Data View**, pe fiecare linie vom introduce câte un caz, adică valorile obținute pe un chestionar. Dacă avem un esantion de 300 de persoane, vom avea 300 de linii. Introducerea fiecărei linii este, fără îndoială, partea cea mai neplăcută din statistică, munca fiind mecanică, obositoare și necesitând mult timp. După introducerea, este bine ca o a doua persoană să verifice datele.

**Introducerea întrebărilor cu răspuns multiplu** este ceva mai problematică. Pentru a putea centraliza o astfel de întrebare, este nevoie ca fiecare variantă de răspuns să fie introdusă ca variabilă separată, care să ia valorile "0"- dacă nu a fost încercuită, și "1"-dacă a fost încercuită. De exemplu, pentru întrebarea *Ce canale media folosiți: a)tv, b)internet, c)radio, d)presă scrisă, e) altele* vom avea 5 variabile de tipul: "utilizarea tv", "utilizarea internet" etc. La final, adunând cazurile cu valoarea "1" pentru fiecare din aceste variabile rezultă numărul de oameni care utilizează fiecare canal. Se poate citi: *89% din populație folosește televiziunea, 68% din populație folosește internetul* etc. Desigur, dacă adunăm procente, în aceste cazuri, rezultatul va fi mai mare de 100%.

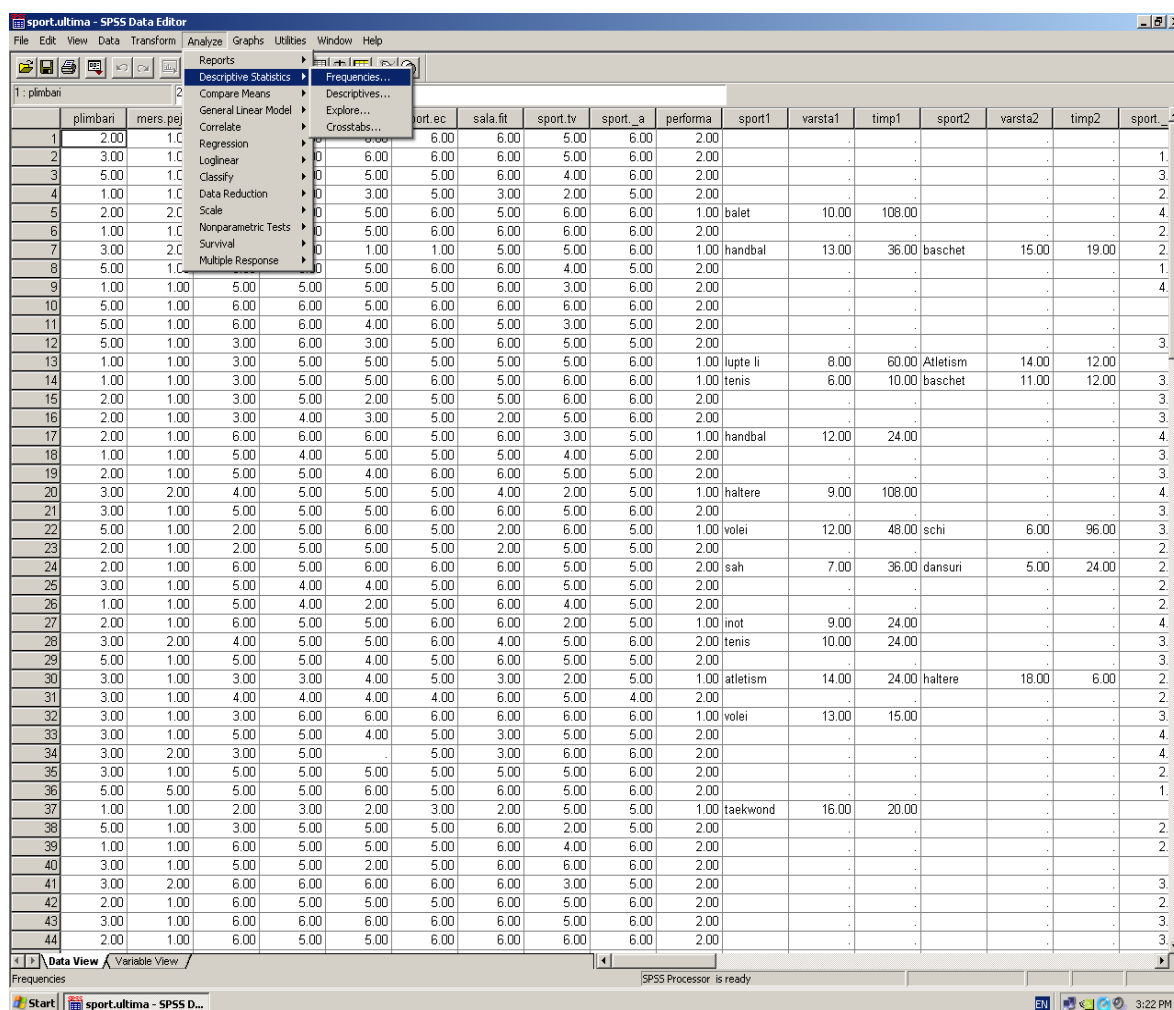
**Introducerea întrebărilor cu răspuns deschis** este și ea problematică. De cele mai multe ori, în realizarea cercetării evităm să avem un număr mare de întrebări deschise pentru că acestea se prelucrează mai greu. La întrebările deschise, variabila va fi de tip *String*, iar cercetătorul va face un rezumat al răspunsului, concentrându-se pe cuvinte cheie.

Ulterior, prelucrarea acestor întrebări se poate face astfel: răspunsurile sunt analizate și cele oarecum similare sunt grupate a posteriori într-un sistem de variante de răspuns, operațiune numită postcodificare. Este și aceasta o operațiune dificilă și greoaie, mai ales dacă varietatea răspunsurilor este mare. După postcodificare, întrebarea deschisă devine o variabilă nominală obișnuită. Însă pentru că ne interesează și nuanțele fiecărui răspuns în parte, este bine să se realizeze și o analiză narativă a răspunsurilor, adică să se "povestească" rezultatele.

## Statistică descriptivă

Pentru a defini caracteristicile grupului, indicatorii statistici cel mai des utilizați sunt:

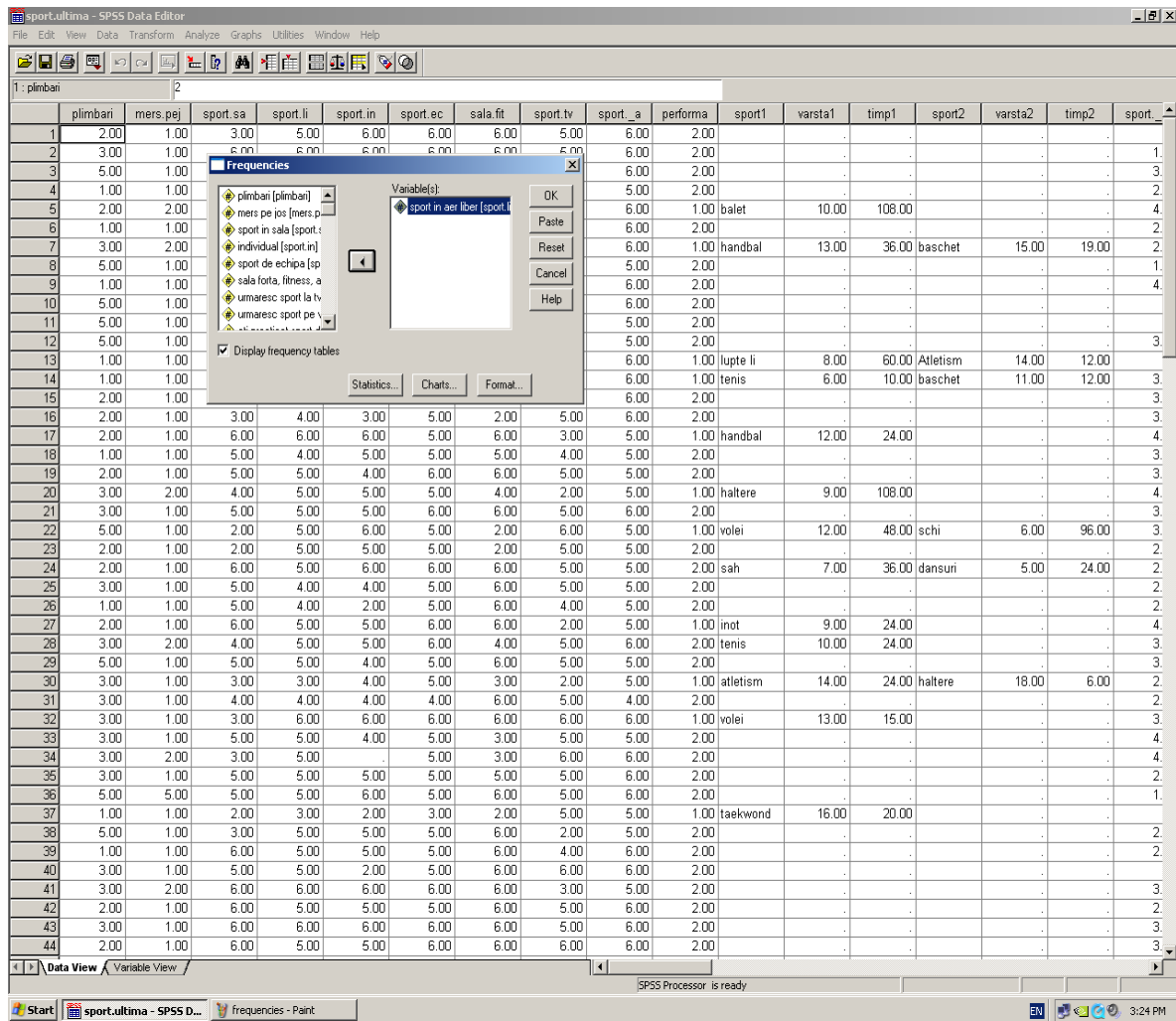
**Frecvență absolută.** Numără la nivel absolut cazurile care se încadrează într-o anumită valoare pe care o ia variabila. Exemplu: pe variabila "gen", care poate lua două valori – "1-masculin", "2-feminin" - frecvența absolută ne arată de câte ori apare valoarea 1, respectiv valoarea 2, deci câți bărbați, respectiv femei avem între subiecți.



The screenshot shows the SPSS Data Editor window with a data table. The menu 'Analyze' is open, and 'Frequencies...' is selected. The data table has the following columns: plimbări, mers.pe, sport.ec, sala.fit, sport.tv, sport\_a, performa, sport1, varsta1, timp1, sport2, varsta2, timp2, sport. The data rows are numbered 1 to 44.

	plimbări	mers.pe	sport.ec	sala.fit	sport.tv	sport_a	performa	sport1	varsta1	timp1	sport2	varsta2	timp2	sport			
1	2.00	1.0	6.00	6.00	5.00	6.00	2.00										
2	3.00	1.0	6.00	6.00	5.00	6.00	2.00							1.			
3	5.00	1.0	5.00	5.00	6.00	4.00	6.00	2.00						3.			
4	1.00	1.0	3.00	5.00	3.00	2.00	5.00	2.00						2.			
5	2.00	2.0	5.00	6.00	5.00	6.00	6.00	1.00	balet	108.00				4.			
6	1.00	1.0	5.00	6.00	6.00	6.00	6.00	2.00						2.			
7	3.00	2.0	1.00	1.00	5.00	5.00	6.00	1.00	handbal	13.00	36.00	baschet	15.00	19.00	2.		
8	5.00	1.0	5.00	6.00	6.00	4.00	5.00	2.00						1.			
9	1.00	1.00	5.00	5.00	5.00	6.00	3.00	6.00	2.00					4.			
10	5.00	1.00	6.00	6.00	5.00	6.00	6.00	6.00	2.00					1.			
11	5.00	1.00	6.00	6.00	4.00	6.00	5.00	3.00	2.00					4.			
12	5.00	1.00	3.00	6.00	3.00	5.00	6.00	5.00	2.00					3.			
13	1.00	1.00	3.00	5.00	5.00	5.00	5.00	6.00	1.00	lupte li	8.00	60.00	Atletism	14.00	12.00	3.	
14	1.00	1.00	3.00	5.00	5.00	6.00	6.00	1.00	tenis	6.00	10.00	baschet	11.00	12.00	3.		
15	2.00	1.00	3.00	5.00	2.00	5.00	5.00	6.00	2.00					3.			
16	2.00	1.00	3.00	4.00	3.00	5.00	2.00	5.00	6.00	2.00				3.			
17	2.00	1.00	6.00	6.00	6.00	5.00	6.00	3.00	5.00	1.00	handbal	12.00	24.00		4.		
18	1.00	1.00	5.00	4.00	5.00	5.00	4.00	5.00	2.00					3.			
19	2.00	1.00	5.00	5.00	4.00	6.00	6.00	5.00	2.00					3.			
20	3.00	2.00	4.00	5.00	5.00	5.00	4.00	2.00	5.00	1.00	haltere	9.00	108.00		4.		
21	3.00	1.00	5.00	5.00	6.00	6.00	5.00	6.00	2.00					3.			
22	5.00	1.00	2.00	5.00	6.00	5.00	2.00	6.00	5.00	1.00	volei	12.00	48.00	schi	6.00	96.00	3.
23	2.00	1.00	2.00	5.00	5.00	5.00	2.00	5.00	2.00					2.			
24	2.00	1.00	6.00	5.00	6.00	6.00	6.00	5.00	5.00	2.00	sah	7.00	36.00	dansuri	5.00	24.00	2.
25	3.00	1.00	5.00	4.00	4.00	5.00	6.00	5.00	5.00	2.00				2.			
26	1.00	1.00	5.00	4.00	2.00	5.00	6.00	4.00	5.00	2.00				2.			
27	2.00	1.00	6.00	5.00	5.00	6.00	6.00	2.00	5.00	1.00	inot	9.00	24.00		4.		
28	3.00	2.00	4.00	5.00	5.00	6.00	4.00	5.00	6.00	2.00	tenis	10.00	24.00		3.		
29	5.00	1.00	5.00	5.00	4.00	5.00	6.00	5.00	5.00	2.00				3.			
30	3.00	1.00	3.00	3.00	4.00	5.00	3.00	2.00	5.00	1.00	atletism	14.00	24.00	haltere	18.00	6.00	2.
31	3.00	1.00	4.00	4.00	4.00	4.00	6.00	5.00	4.00	2.00				2.			
32	3.00	1.00	3.00	6.00	6.00	6.00	6.00	6.00	1.00	volei	13.00	15.00		3.			
33	3.00	1.00	5.00	5.00	4.00	5.00	3.00	5.00	2.00					4.			
34	3.00	2.00	3.00	5.00		5.00	3.00	6.00	6.00	2.00				4.			
35	3.00	1.00	5.00	5.00	5.00	5.00	5.00	6.00	6.00	2.00				2.			
36	5.00	5.00	5.00	5.00	6.00	5.00	6.00	5.00	6.00	2.00				1.			
37	1.00	1.00	2.00	3.00	2.00	3.00	2.00	5.00	5.00	1.00	taekwond	16.00	20.00		2.		
38	5.00	1.00	3.00	5.00	5.00	5.00	6.00	2.00	5.00	2.00				2.			
39	1.00	1.00	6.00	5.00	5.00	5.00	6.00	4.00	6.00	2.00				2.			
40	3.00	1.00	5.00	5.00	2.00	5.00	6.00	6.00	2.00					2.			
41	3.00	2.00	6.00	6.00	6.00	6.00	3.00	5.00	2.00					3.			
42	2.00	1.00	6.00	5.00	5.00	5.00	6.00	6.00	2.00					2.			
43	3.00	1.00	6.00	6.00	6.00	6.00	6.00	5.00	6.00	2.00				3.			
44	2.00	1.00	6.00	5.00	5.00	6.00	6.00	2.00						3.			

**Frecvență relativă.** Este frecvența unei valori raportată la numărul total de cazuri. Mai exact, este frecvența calculată în procente. Programul SPSS furnizează două tipuri de frecvențe relative. În cazul unor valori lipsă (non-răspunsuri), SPSS calculează diferențiat *frecvența relativă validă*, care este raportată nu la numărul total de cazuri ci la numărul de cazuri minus valorile lipsă.



Rezultatele în urma acestui tip de comandă în SPSS apar sub forma unui tabel de frecvențe, care arată astfel. Observați diferența dintre frecvența absolută, cea procentuală și cea procentuală validă.

**sport in aer liber**

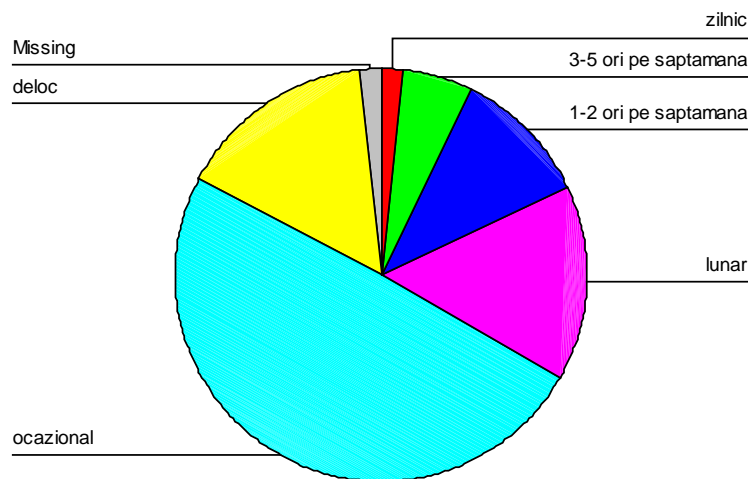
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	zilnic	2	1.8	1.8	1.8
	3-5 ori pe saptamana	6	5.4	5.5	7.3
	1-2 ori pe saptamana	12	10.8	11.0	18.3
	lunar	17	15.3	15.6	33.9
	ocazional	55	49.5	50.5	84.4
	deloc	17	15.3	15.6	100.0
	Total	109	98.2	100.0	
Missing	System	2	1.8		
Total		111	100.0		

Diferența între ultimele două este dată de faptul că avem 2 valori lipsă (missing). De asemenea, remarcați coloana de procente cumulate (*cumulative percent*), care are relevanță

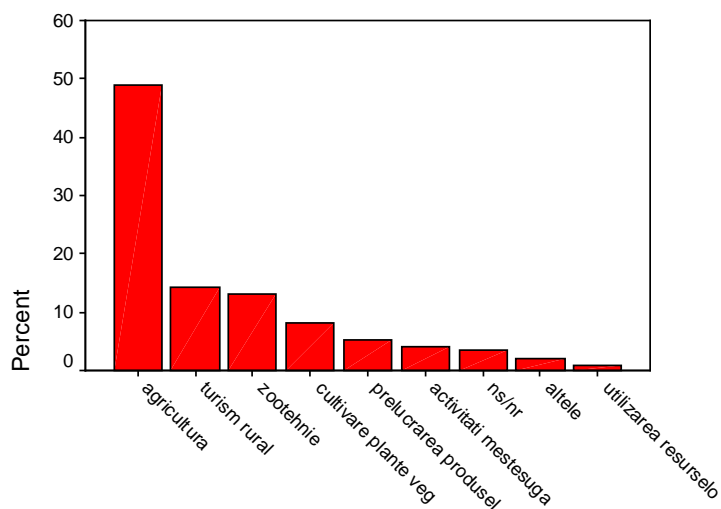
doar pentru variabilele cel puțin ordinale (deci nu pentru cele nominale). Interpretarea procentului cumulat pe exemplul dat, se face în felul următor: 33,9 % din populație practică sport în aer liber lunar sau mai des, 18,3% din populație practică sport în aer liber cel puțin de 1-2 ori pe săptămână.

Pentru frecvențe, se pot realiza și grafice astfel: comanda **Analyze – Descriptive Statistics – Frequencies** deschide fereastra pentru frecvențe. Aici se poate apăsa **Charts**, unde există trei opțiuni de grafic: *Pie*, *Histogram* și *Bar Chart*. Exemplu de grafice apar mai jos.

sport in aer liber



domenii de activitate 1



domenii de activitate 1



De remarcat este faptul că, și în funcție de varianta de SPSS, graficele în SPSS pot să fie mai ”arătoase” sau mai puțin estetice. Grafice se pot face, însă, pe aceleași date, luate cu **Copy – Paste** și utilizate în Microsoft Excel, de unde se utilizează opțiunea **Pivot Table** și **Charts**, iar variabilele sunt plasate pe tabele cu *drag and drop*.

**Observatie:** orice grafic sau tabel realizat de SPSS va fi plasat într-un **Output** separat. Din acest output, datele pot fi copiate în Word sau Powerpoint, dar trebuie ca la copiere să se aleagă opțiunea **Copy object**, în loc de simplul **Copy**.

### **Indicatori ai tendinței centrale**

**Medie.** Valorile din sir sunt adunate și împărțite la numărul de cazuri valide.

**Mediana.** Valoarea centrală în ordine crescătoare a valorilor înregistrate. Practic, mediana reprezintă valoarea de mijloc, astfel încât sirul se împarte în două jumătăți egale: prima jumătate sub valoarea medianei, a doua jumătate peste.

**Valoarea modală** sau **modul.** Valoarea cel mai des întâlnită în sir. Poate fi determinată și din tabelul de frecvențe.

### **Indicatori ai împrăstierii datelor**

**Amplitudinea.** Diferența dintre valoarea maximă și valoarea minimă din sir.

**Abaterea standard.** Pentru că uneori media nu este suficientă pentru a înțelege felul în care sunt distribuite valorile, se utilizează deseori acest indicator, care reprezintă media diferențelor dintre fiecare valoare și medie. Abaterea standard este un indicator util pentru că ne spune de fapt cât de omogen este sirul de valori.

Să luăm următorul exemplu: alegem variabila ”tolerantă religioasă” care stabilim că va lua valori de la 1 la 10 pentru subiecții din două grupuri diferite. La sfârșitul culegerii datelor calculăm media și remarcăm că grupul A are aceeași medie a toleranței cu grupul B: 7. Am

putea concluziona, deci, că grupurile sunt identice din punct de vedere al toleranței religioase pe care o declară. Însă, dacă ne uităm mai atent, valorile toleranței pentru fiecare membru al grupului A sunt: 7, 7, 7, 7, 7, 7, 7, 7, 7, deci o perfectă omogenitate a atitudinilor, iar pentru grupul B valorile sunt: 4, 10, 10, 10, 4, 4, 10, 4, deci o factionalizare extremă a atitudinilor, în care avem mulți indivizi foarte puțin toleranți pe de o parte și la fel de mulți indivizi foarte toleranți pe de alta, deci o situație eterogenă, instabilă și potențial problematică. Deși media pentru cele două situații este aceeași, abaterea standard pentru situația A este 0 (situație aproape imposibil de întâlnit într-un exemplu realist), iar abaterea standard pentru B este 3.

Ne dăm seama mai bine dacă o abatere este mare sau mică dacă raportăm rezultatul la indicatorul anterior: *amplitudinea*.

**Coeficientul de variabilitate.** Raportarea abaterii standard la media sirului, în procente. Pe exemplul anterior, B, unde abaterea este 3, iar media este 7, coeficientul de variabilitate  $3 \cdot 100 / 7 = 42,8\%$ , coeficient cu o valoare mare.

Indicatorii tendinței centrale și cei ai împrăstierii datelor se calculează în SPSS din fereastra de frecvențe, apăsând **Statistics** și selectând fiecare indicator în parte, dacă este de interes.

## Statistică analitică

Testarea de ipoteze într-o cercetare presupune mai mult decât simpla descriere a datelor. Trebuie realizate corelații între variabilele măsurate. Cea mai simplă modalitate de a corela două variabile este aceea de a le plasa într-un tabel încrucișat (crosstabs). Operațiunea este evidențiată mai jos:

The screenshot shows the SPSS Data Editor interface. A data table is visible with columns for various variables like 'plimbări', 'mers.pej', 'sport.sa', etc. A 'Crosstabs' dialog box is open, showing a list of variables on the left and a list of variables in the 'Row(s)' and 'Column(s)' fields. The 'Cells...' button is highlighted, indicating the next step in the process.

	plimbări	mers.pej	sport.sa	sport.li	sport.in	sport.ec	sala.fit	sport.tv	sport_a	performa	sport1	varsta1	timp1	sport2	varsta2	timp2	sport...
1	2.00	1.00	3.00	5.00	6.00	6.00	6.00	5.00	6.00	2.00							1.
2	3.00							5.00	6.00	2.00							3.
3	5.00							4.00	6.00	2.00							2.
4	1.00							2.00	5.00	2.00							4.
5	2.00							6.00	6.00	1.00	balet	10.00	108.00				2.
6	1.00							6.00	6.00	2.00							2.
7	3.00							5.00	6.00	1.00	handbal	13.00	36.00	baschet	15.00	19.00	2.
8	5.00							4.00	5.00	2.00							1.
9	1.00							3.00	6.00	2.00							4.
10	5.00							6.00	6.00	2.00							
11	5.00							3.00	5.00	2.00							
12	5.00							5.00	5.00	2.00							3.
13	1.00							5.00	6.00	1.00	lupte li	8.00	60.00	Atletism	14.00	12.00	
14	1.00							6.00	6.00	1.00	tenis	6.00	10.00	baschet	11.00	12.00	3.
15	2.00							6.00	6.00	2.00							3.
16	2.00							5.00	6.00	2.00							3.
17	2.00							3.00	5.00	1.00	handbal	12.00	24.00				4.
18	1.00							4.00	5.00	2.00							3.
19	2.00							5.00	5.00	2.00							3.
20	3.00							2.00	5.00	1.00	haltere	9.00	108.00				4.
21	3.00	1.00	5.00	5.00	5.00	6.00	6.00	5.00	6.00	2.00							3.
22	5.00	1.00	2.00	5.00	6.00	5.00	2.00	6.00	5.00	1.00	volei	12.00	48.00	schii	6.00	96.00	3.
23	2.00	1.00	2.00	5.00	5.00	5.00	2.00	5.00	5.00	2.00							2.
24	2.00	1.00	6.00	5.00	6.00	6.00	6.00	5.00	5.00	2.00	sah	7.00	36.00	dansuri	5.00	24.00	2.
25	3.00	1.00	5.00	4.00	4.00	5.00	6.00	5.00	5.00	2.00							2.
26	1.00	1.00	5.00	4.00	2.00	5.00	6.00	4.00	5.00	2.00							2.
27	2.00	1.00	6.00	5.00	5.00	6.00	6.00	2.00	5.00	1.00	inot	9.00	24.00				4.
28	3.00	2.00	4.00	5.00	5.00	6.00	4.00	5.00	6.00	2.00	tenis	10.00	24.00				3.
29	5.00	1.00	5.00	5.00	4.00	5.00	6.00	5.00	5.00	2.00							3.
30	3.00	1.00	3.00	3.00	4.00	5.00	3.00	2.00	5.00	1.00	atletism	14.00	24.00	haltere	18.00	6.00	2.
31	3.00	1.00	4.00	4.00	4.00	4.00	6.00	5.00	4.00	2.00							2.
32	3.00	1.00	3.00	6.00	6.00	6.00	6.00	6.00	6.00	1.00	volei	13.00	15.00				3.
33	3.00	1.00	5.00	5.00	4.00	5.00	3.00	5.00	5.00	2.00							4.
34	3.00	2.00	3.00	5.00		5.00	3.00	6.00	6.00	2.00							4.
35	3.00	1.00	5.00	5.00	5.00	5.00	5.00	5.00	6.00	2.00							2.
36	5.00	5.00	5.00	5.00	6.00	5.00	6.00	5.00	6.00	2.00							1.
37	1.00	1.00	2.00	3.00	2.00	3.00	2.00	5.00	5.00	1.00	haekwond	16.00	20.00				2.
38	5.00	1.00	3.00	5.00	5.00	6.00	5.00	6.00	2.00	5.00	2.00						2.
39	1.00	1.00	6.00	5.00	5.00	5.00	6.00	4.00	6.00	2.00							2.
40	3.00	1.00	5.00	5.00	2.00	5.00	6.00	6.00	6.00	2.00							
41	3.00	2.00	6.00	6.00	6.00	6.00	6.00	3.00	5.00	2.00							3.
42	2.00	1.00	6.00	5.00	5.00	5.00	6.00	5.00	6.00	2.00							2.
43	3.00	1.00	6.00	6.00	6.00	6.00	6.00	5.00	6.00	2.00							3.
44	2.00	1.00	6.00	5.00	5.00	6.00	6.00	6.00	6.00	2.00							3.

De pe butonul *Cells* care apare în fereastra *Crosstabs*, se poate selecta evidențierea frecvențelor compuse variabile, adică afișarea de procente atât pe coloane, cât și pe rânduri. Rezultatul unei crosstabulări simple este afișat mai jos:



Se observă pe exemplul dat, dacă ne uităm la procente, că Jucu de Sus și Visea au probleme destul de mari cu drumurile.

- Gadalin și Jucu de Mijloc par cel mai afectate de transportul în comun
- Jucu de Sus are probleme cu modernizarea școlii și a grădinitei
- Gadalin și Visea mai afectate de problema locurilor de muncă
- Gadalin, Visea și Jucu de Mijloc arată că au problema serviciilor medicale

Dacă dorim să testăm și forța asocierii între două variabile, se poate calcula un coeficient de corelație. Un astfel de coeficient verifică în ce măsură cele două variabile variază împreună. O creștere a variabilei X se corelează cu o creștere sau cu o scădere a variabilei Y? Cât de puternică este variația? Putem afla răspunsul la aceste întrebări, însă fără să facem distincția între cauză și efect, prin indici Pearson sau Kendall. Acești indici pot lua valori între -1 și +1. Cu cât sunt mai aproape de 0, acești coeficienți demonstrează o slabă corelație între variabilele testate. Dacă sunt apropiați de -1, coeficienții arată o corelare inversă (cu cât crește X scade Y), dacă sunt apropiați de +1, arată o corelare pozitivă (variabilele cresc împreună). De remarcat că variabilele nominale, nefiind ordonate natural, nu pot fi corelate decât cel mult dacă sunt dihotomice (de exemplu: "da" – "nu").

SPSS ne semnalează, în același timp, dacă rezultatul corelării este semnificativ din punct de vedere statistic, prin calcularea coeficientului Sig. Dacă acesta este sub 0.05, corelația este semnificativă și se contrazice ipoteza nulă (adică ipoteza conform căreia nu există legătură între variabile). Vedem mai jos un tabel de corelație între două variabile, cu calcularea indicelui Pearson:

**Correlations**

		sport in aer liber	probleme de sanatate
sport in aer liber	Pearson Correlation	1.000	-.158
	Sig. (2-tailed)	.	.102
	N	109	108
probleme de sanatate	Pearson Correlation	-.158	1.000
	Sig. (2-tailed)	.102	.
	N	108	110

În situația prezentată, coeficientul Pearson este -0.158, adică o slabă corelație între frecvența problemelor de sănătate și frecvența practicării sporturilor în aer liber. Practic nu există legătură între cele două variabile. Coeficientul Sig. nu este în limitele căutate de noi. Atunci

când corelata este semnificativă, Sig. din tabel este cel puțin sub 0.05, iar în notațiile pe care le folosim la interpretare vom scrie *Cele două variabile sunt corelate la nivel -0.158, p <0.05*. Când Sig. este în intervalul relevant, SPSS ne ajută prin semnalarea corelației cu semnul \*.

Iată mai jos un exemplu pe corelarea scorurilor obținute de un esanțion de elevi la diferite secțiuni ale testării. Aici au fost corelate simultan 5 variabile diferite, dintre care și una nominală dihotomică (genul). S-au descoperit un număr de 7 corelații semnificative la nivel  $p < 0.01$  (grad de încredere a corelației foarte mare), unde Sig. este chiar cu valoarea 0.000.

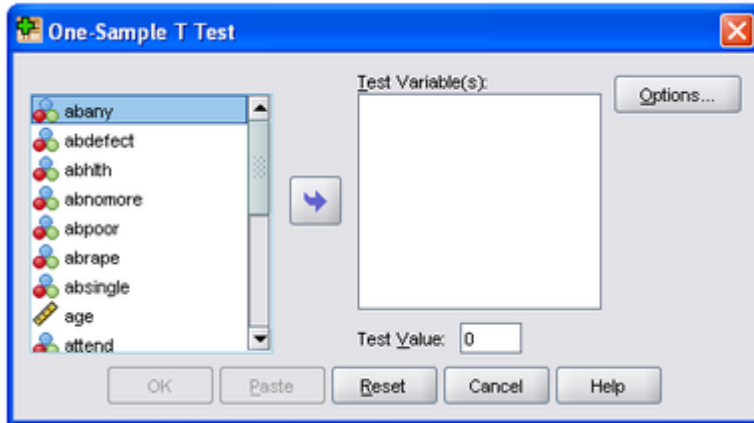
**Correlations**

		reading score	writing score	math score	science score	female
reading score	Pearson Correlation <sup>a</sup>	1	.597**	.662**	.630**	-.053
	Sig. (2-tailed) <sup>b</sup>	.	.000	.000	.000	.455
	N <sup>c</sup>	200	200	200	200	200
writing score	Pearson Correlation	.597**	1	.617**	.570**	.256**
	Sig. (2-tailed)	.000	.	.000	.000	.000
	N	200	200	200	200	200
math score	Pearson Correlation	.662**	.617**	1	.631**	-.029
	Sig. (2-tailed)	.000	.000	.	.000	.680
	N	200	200	200	200	200
science score	Pearson Correlation	.630**	.570**	.631**	1	-.128
	Sig. (2-tailed)	.000	.000	.000	.	.071
	N	200	200	200	200	200
female	Pearson Correlation	-.053	.256**	-.029	-.128	1
	Sig. (2-tailed)	.455	.000	.680	.071	.
	N	200	200	200	200	200

\*\* Correlation is significant at the 0.01 level (2-tailed).

Deci, de pe tabelul de mai sus putem citi: *scorul pentru scris este corelat la nivel 0.597 cu scorul obținut de subiecți pentru citit,  $p < 0.01$ , ceea ce indică o puternică legătură între cele două abilități. Cu cât elevii citesc mai bine, cu atât scriu mai bine și invers*. Este important să nu uităm că o corelație merge în ambele sensuri. Ea nu verifică efectul și cauza, însă uneori cauza și efectul pot fi determinate prin raționament logic. Putem citi de asemenea că *femeile obțin scoruri semnificativ mai bune decât bărbații la capitolul scriere, cele 2 variabile fiind corelate cu un coeficient 0.256,  $p < 0.01$* .

O altă variantă de prezentare a legăturii dintre 2 variabile este cea a **comparării mediilor** (atunci când mediile pot fi calculate) sau *testul t*. Comanda este **Analyze – Compare Means – One-Sample T Test**. Se va deschide o fereastră care arată astfel:



La *test value* vom trece media de la una dintre variabilele pe care le comparăm, iar din stânga o vom selecta pe cealaltă. SPSS va compara *media prevăzută/asteptată* cu *media observată* rezultând din nou o asociere mai mult sau mai puțin solidă. Rezultatul arată cam așa:

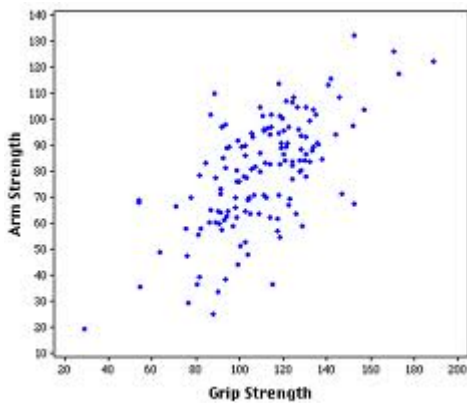
**Paired Samples Test**

		Paired Differences							
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference		t	df	Sig. (2-tailed)
					Lower	Upper			
Pair 1	BI - BI2	.023	1.165	.178	-.335	.382	.131	42	.896

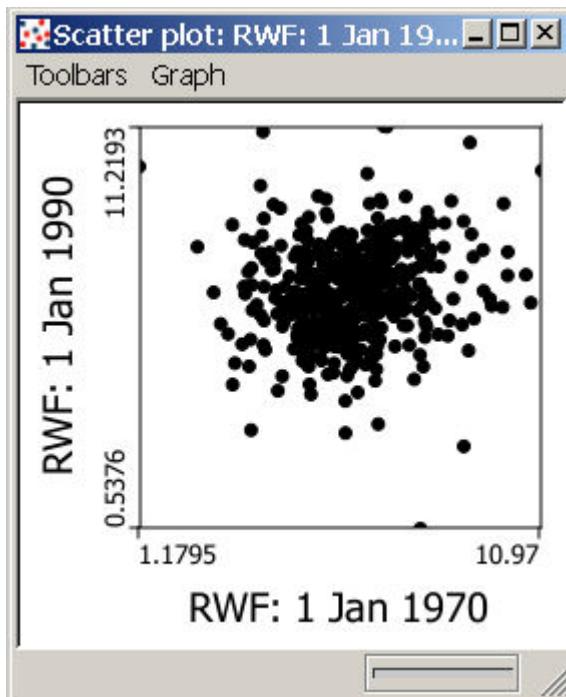
În acest exemplu, nu există o diferență semnificativă, lucru pe care îl observăm din nou verificând Sig., care nu este mai mic de 0.05. Valoarea testului se găsește în coloana "t".

O posibilitate de a prefigura legătura dintre 2 variabile este și aceea de a face un grafic de tip *Scatterplot*, adică norul de puncte, unde punctele sunt intersecția valorilor de pe cele două variabile, pentru fiecare caz. Acest lucru se realizează de la **Graphs – Scatter**, apoi în fereastra deschisă se alege tipul de grafic și se apasă butonul **Define** pentru a selecta variabilele care vor fi asociate. Dacă punctele rezultate sunt relativ grupate în jurul unei drepte imaginare, ascendentă sau descendentă, înseamnă că există o legătură între variabile, care poate fi testată ulterior. Mai jos, exemplul dat arată că, în linii mari, cazurile se distribuie în

jurul unei drepte ascendente, astfel că pe măsură ce crește *arm strength* (*forta bratului*), crește și *grip strength* (*forta prizei*).



Un scatterplot care nu relevă o legătură dintre variabile arată așa:



Există și alte posibile operații de corelare în SPSS. Una des întâlnită este regresia. Aceasta este o operațiune prin care se încearcă explicarea unei variabile (să zicem *rezultatele la învățatură*) prin plasarea ei ca rezultat al unei funcții, în care apar alte variabile cu rol explicativ. Astfel, putem testa în raport cu rezultatele la învățatură dependentă de alte două variabile: *coeficientul de inteligență* și *prezența la cursuri*. Astfel, practic plasăm pe o singură linie (funcția), toate punctele din scatterplot (desigur o putem face doar dacă observăm din grafic că punctele sunt cât de puțin grupate după acea dreaptă imaginară). Dreapta imaginară din scatterplot este dreapta de regresie, care are la bază o formulă. În științele socio-umane,



însă, acest model ultra-matematicizat de lucru este din ce în ce mai puțin acceptat ca relevant științific. Un comportament sau o caracteristică socială nu poate fi explicată corect prin modele exagerat sau artificial matematicizate, ci mai degrabă prin contribuția unor explicații de ordin calitativ la demersul științific.

Datele obținute în urma tuturor operațiunilor prezentate în acest mic ghid sunt folosite pentru explicarea unei varietăți de fenomene, aplicațiile SPSS fiind practic infinite. În marketing și publicitate aceste metode cantitative sunt folosite în ultimul timp mai puțin decât metodele calitative de tipul focus-grupului. Totuși, datele legate de vânzări, audiențe, caracteristici socio-demografice, dar și psihografice ale publicului țintă, sunt tratate ca elemente cantitative și sunt corelate într-o varietate de combinații.

Mai mult, la un pitch în care publicitarul va încerca să-ți convingă clientul să semneze un contract cu el, iar apoi în momentul în care clientul va dori să vadă rezultatele pentru care a plătit, elementele de statistică vor fi cruciale pentru a convinge. Nimeni nu va investi bani într-o acțiune ale cărei rezultate nu sunt cuantificabile și în care nu se poate calcula un *return of investment*.

#### Bibliografie:

Coakes, Sheridan J. (2005). *SPSS. Analysis without anguish*. Wiley and Sons, Australia

Griffith, Arthur (2010). *SPSS for dummies*. Wiley Publishing, Hoboken

Jabă, Elisabeta și Ana Grama (2004). *Analiza statistică cu SPSS sub windows*. Polirom, Iași

Leech, Nancy și Karen Barrett (2005). *SPSS for intermediate statistics*. Lawrence Erlbaum, New Jersey