ANALIZA DATELOR SOCIALE. ÎNDRUMAR DE LABORATOR

Adrian Hatos
Universitatea din Oradea
2010

Cuprins

Recodificări, verificarea recodificării, variabile calculate, numărări de valori 3

Distribuții de frecvență și grafice cu variabile categorice 16

Teste non-parametrice 44

Teste non-parametrice ..... **Error! Bookmark not defined.**

Asocierea. Testul Chi-pătrat 50

Compararea mediilor pe două grupuri independente (testul t) 73

Coeficienţi de fidelitate 78

Recodificări, verificarea recodificării, variabile calculate, numărări de valori

Deschideți baza de date droguri2000.sav pe care o găsiți pe web la adresa:

http://ahatos.rdsor.ro/date/droguri_2000.sav

Tema 1: Recodificarea variabilelor

Definiție: transformarea valorilor unei variabile.

Problema: recodificarea variabilei *a1* în 3 categorii.

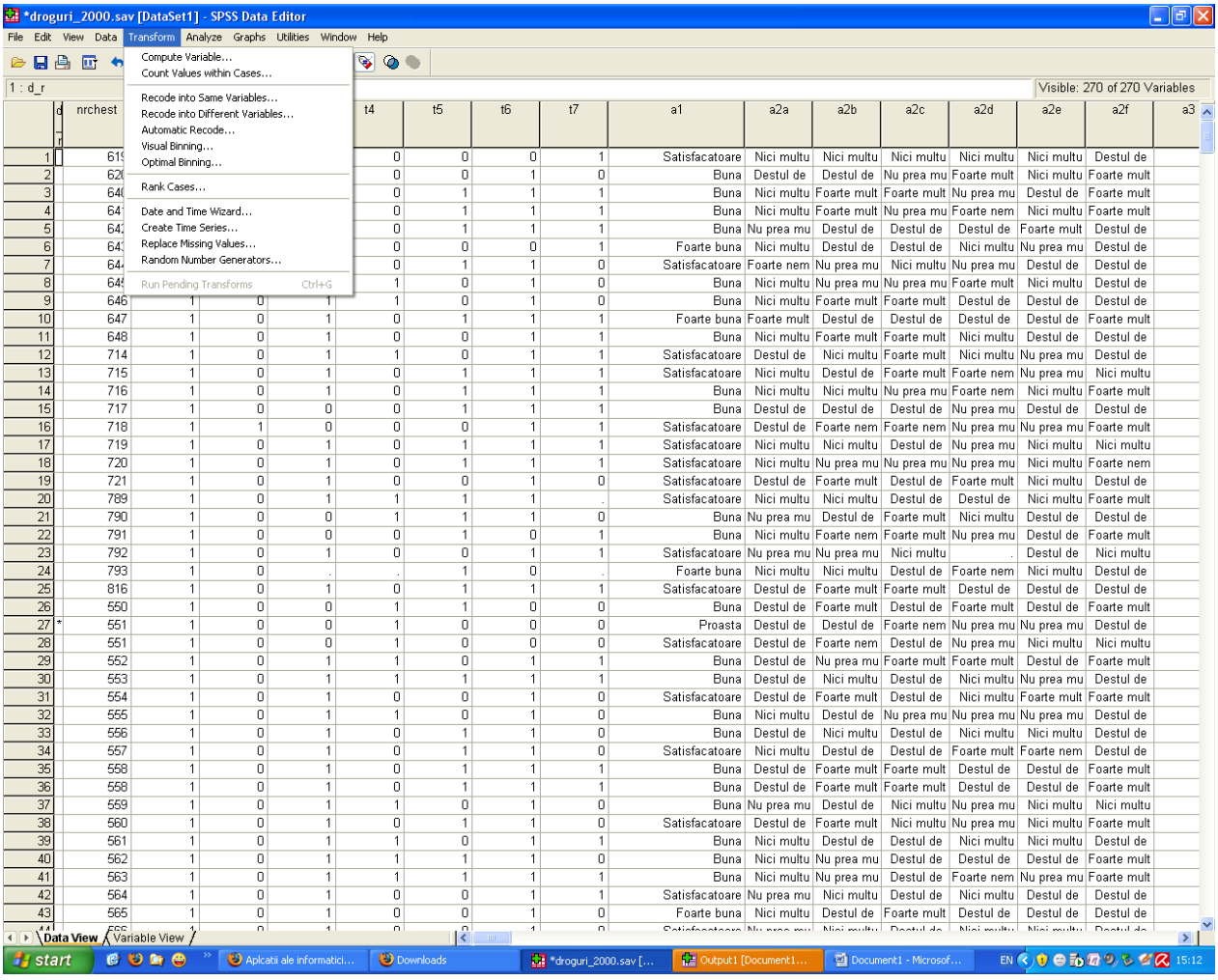
Soluție

Variabila a1, corespunde întrebării: *Cum apreciezi starea sănătății tale?* Variantele de răspuns definite ca valide sunt 5, codificate cu numere: 1- Foarte proasta,2- Proasta, 3- Satisfacatoare,4 - Buna, 5- Foarte buna.

Tema ne solicită transformarea scalei variabilei într-una cu trei categorii. Corespondența dintre valorile vechi şi cele noi este:

Valori vechi	Valori noi
1. Foarte proastă	1
2. Proastă	
3. Satisfăcătoare	2
4. Bună	3
5. Foarte bună	

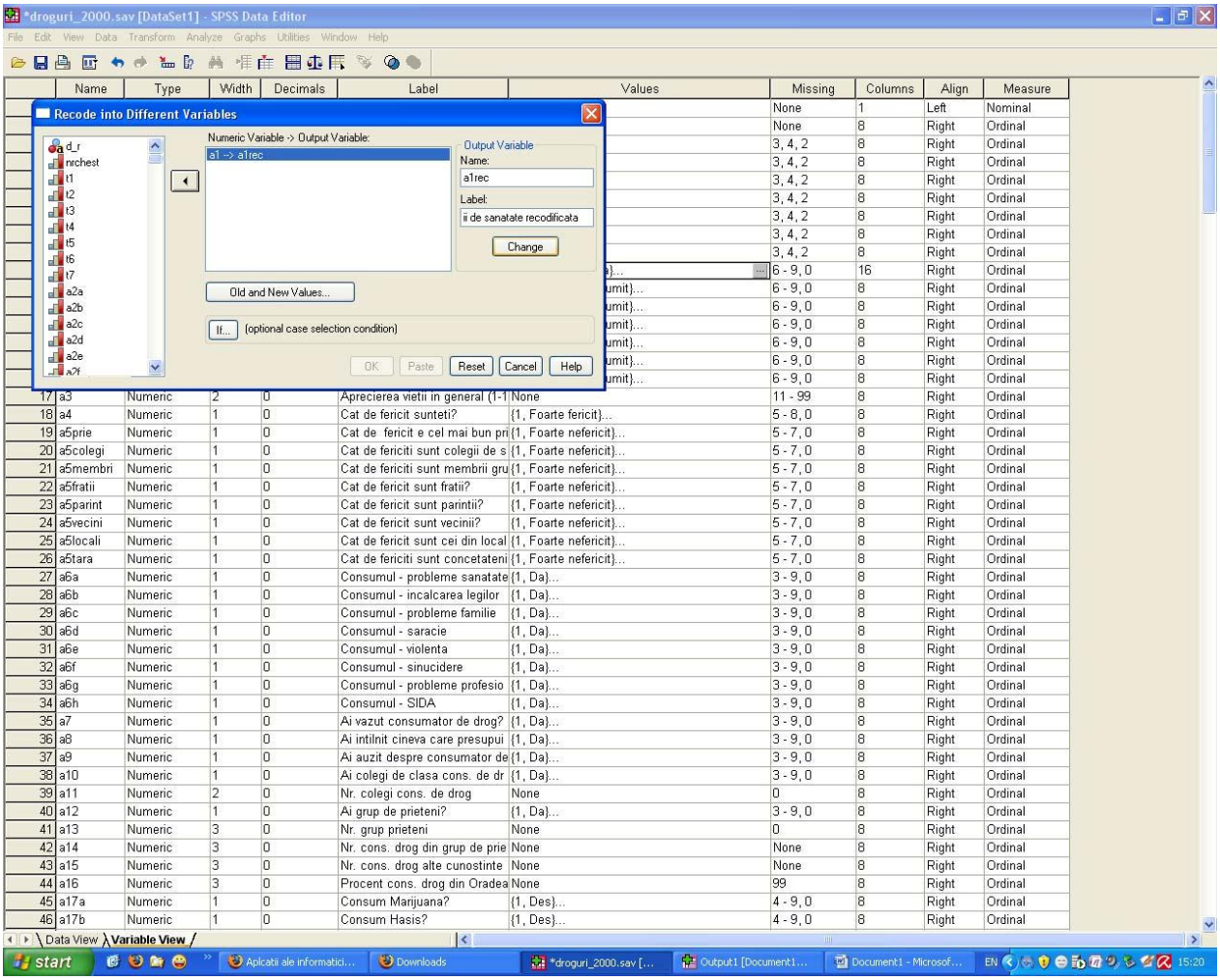
Deschideți opțiunea Transform din meniul principal. La opțiunea recode aveți două variante: Recode into the same variable şi Recode into Diffrent Variable. Întrucât recodifcarea presupune pierdere de informație, vă recomand să alegeți opțiunea a doua, care presupune generarea unei noi variabile.



1. În fereastra de dialog care s-a deschi introducem, pe rând:

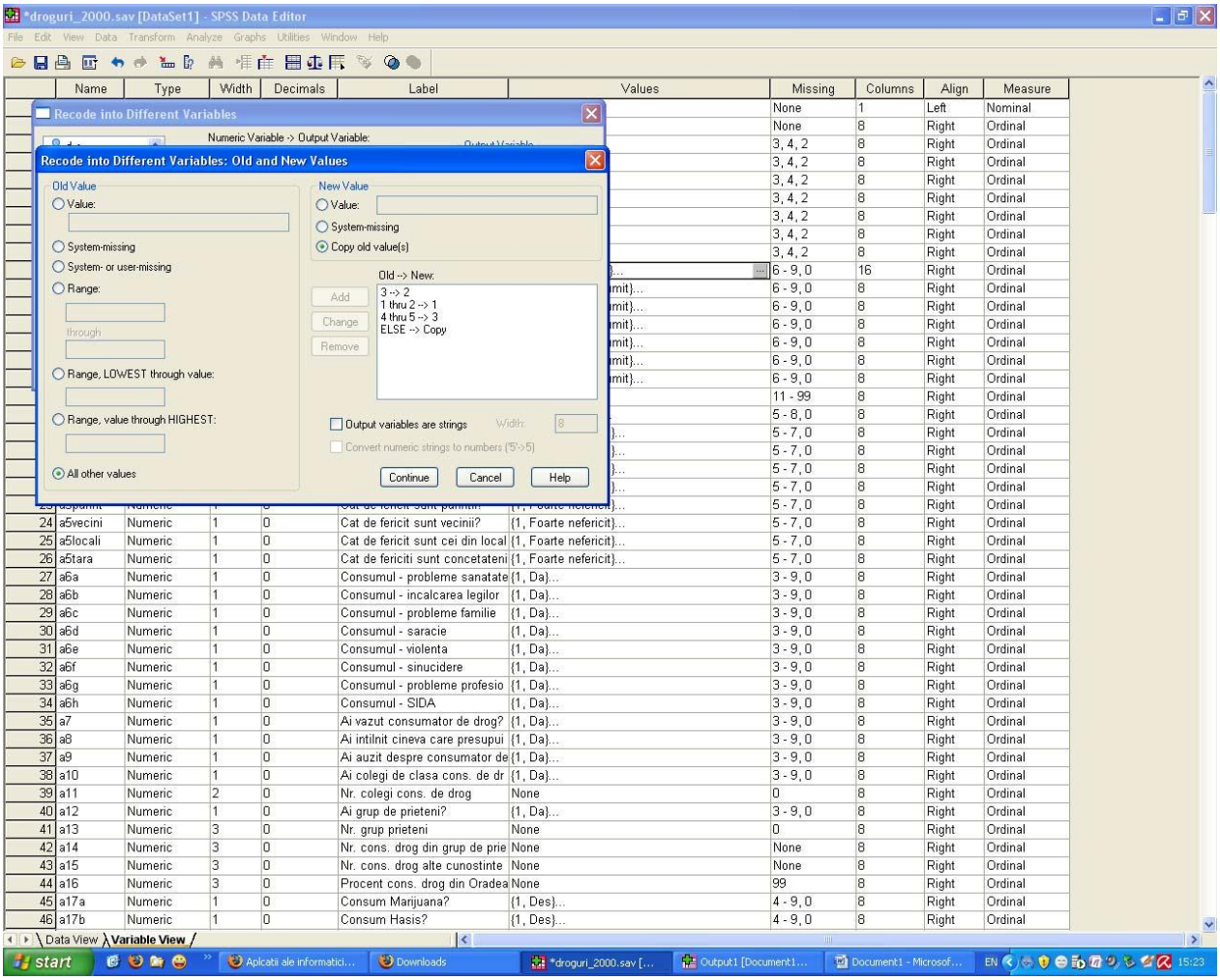
- Variabila care este recodificată (a1)
- Numele noii variabile, rezultată prin recodificare (a1rec)
- Eticheta noii variabile

2. Apoi apăsăm butonul **Change**
3. Apăsăm butonul **Old and New Values**



În noua fereastră de dialog definim corespondența dintre valorile din variabila originală şi variabila nouă.

1. Range: 1 thru 2 -Æ 1
2. 3Æ 2
3. Range: 4 thru 5 Æ3
4. All other valuesÆ Copy old values (pentru restul valorilor, ex. 0, 8, 9, care nu fac parte din scala de atitudini)
5. Apăsăm butonul **Continue**



La capătul bazei de date a fost scrisă noua variabilă, a1rec, ceea ce se poate verifica în Data view sau în Variable view.

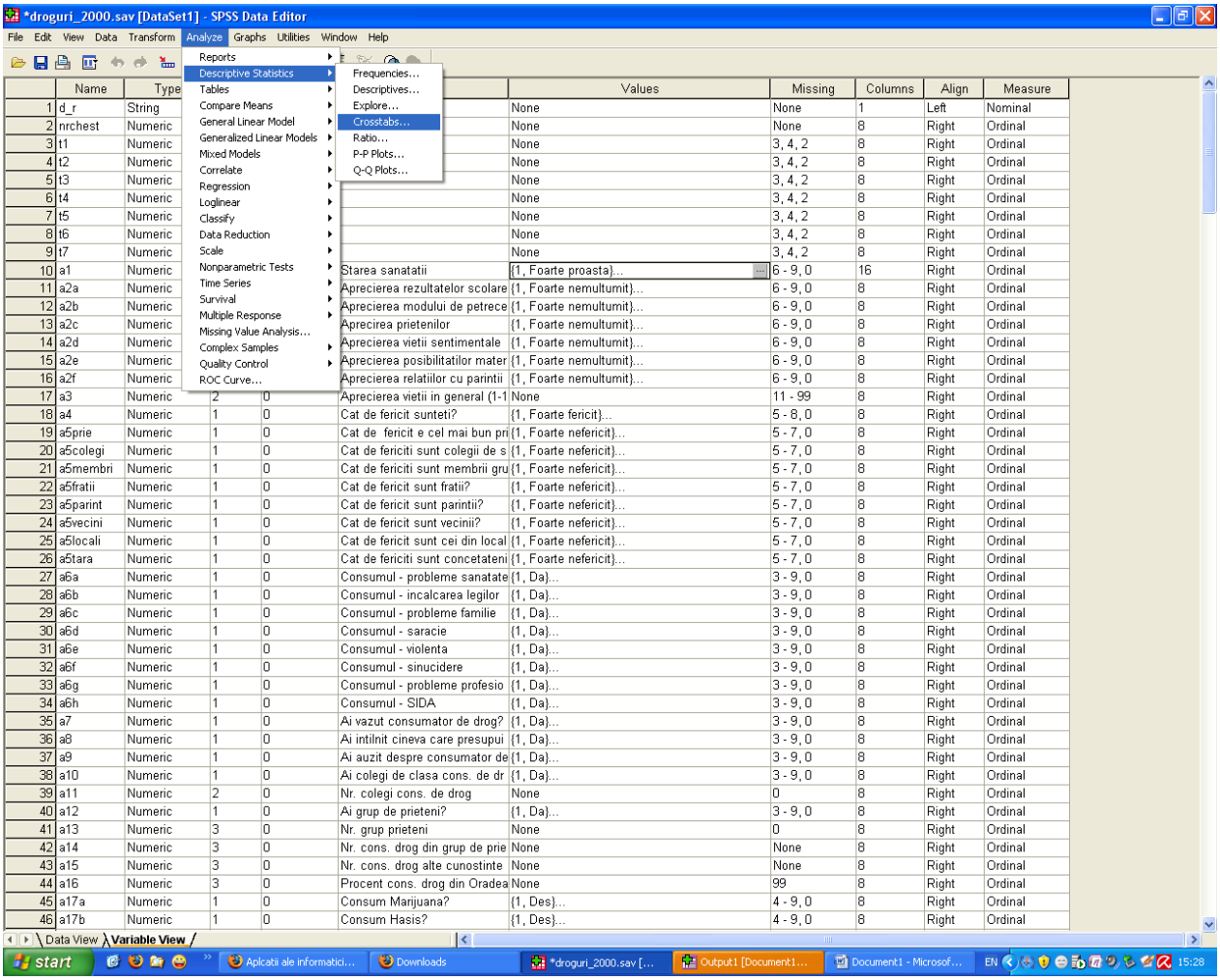
Tema 2: verificarea recodificării – realizarea de tabele încrucişate simple

Problema: verificarea recodificării anterioare

Soluţie

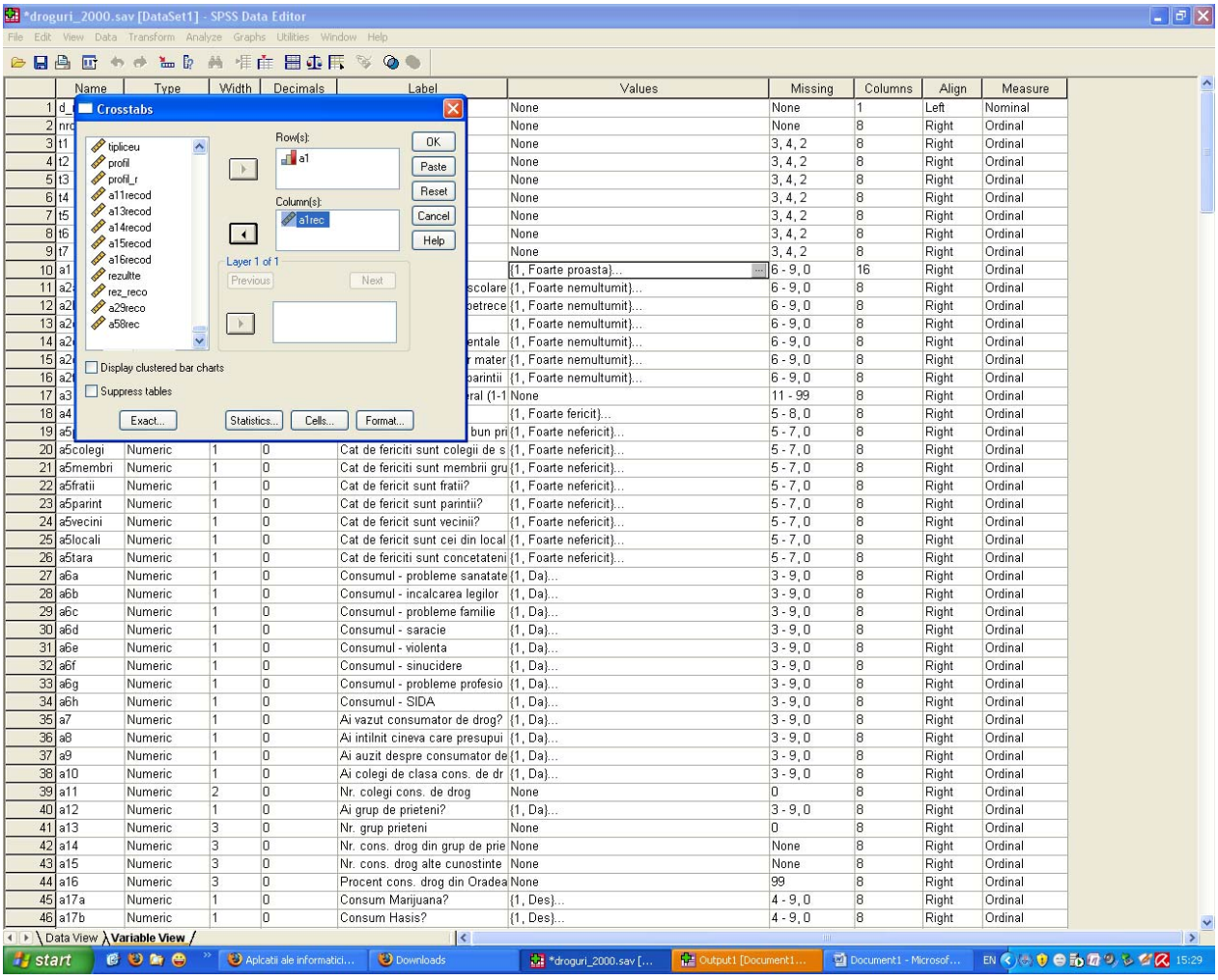
Verificarea recodificării se poate face precis, în acest caz, cu prin construirea unui tabel încrucişat al variabilei originale şi a celei noi.

1. Din meniul principal alegeţi opţiune Analyze Æ Descriptive Statistics Æ Crosstabs

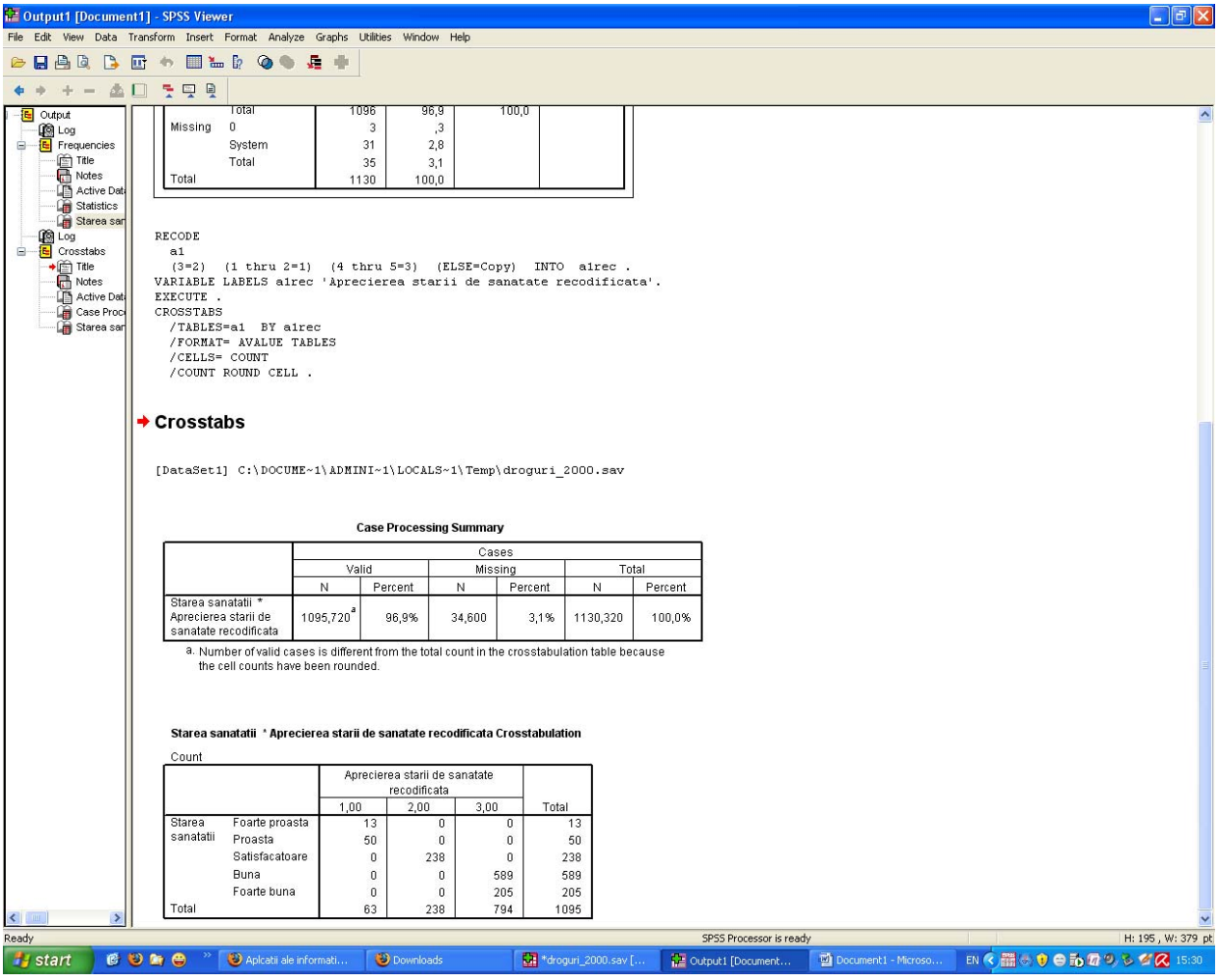


	Name	Type				Values	Missing	Columns	Align	Measure
1	d_r	String				None	None	1	Left	Nominal
2	nrchest	Numeric				None	None	8	Right	Ordinal
3	t1	Numeric				None	3, 4, 2	8	Right	Ordinal
4	t2	Numeric				None	3, 4, 2	8	Right	Ordinal
5	t3	Numeric				None	3, 4, 2	8	Right	Ordinal
6	t4	Numeric				None	3, 4, 2	8	Right	Ordinal
7	t5	Numeric				None	3, 4, 2	8	Right	Ordinal
8	t6	Numeric				None	3, 4, 2	8	Right	Ordinal
9	t7	Numeric				None	3, 4, 2	8	Right	Ordinal
10	a1	Numeric			Starea sanatatii	{1, Foarte proasta}...	6 - 9, 0	16	Right	Ordinal
11	a2a	Numeric			Aprecierea rezultatelor scolare	{1, Foarte nemultumit}...	6 - 9, 0	8	Right	Ordinal
12	a2b	Numeric			Aprecierea modului de petrece	{1, Foarte nemultumit}...	6 - 9, 0	8	Right	Ordinal
13	a2c	Numeric			Aprecirea prietenilor	{1, Foarte nemultumit}...	6 - 9, 0	8	Right	Ordinal
14	a2d	Numeric			Aprecierea vietii sentimentale	{1, Foarte nemultumit}...	6 - 9, 0	8	Right	Ordinal
15	a2e	Numeric			Aprecierea posibilitatilor mater	{1, Foarte nemultumit}...	6 - 9, 0	8	Right	Ordinal
16	a2f	Numeric			Aprecierea relatiilor cu parintii	{1, Foarte nemultumit}...	6 - 9, 0	8	Right	Ordinal
17	a3	Numeric	2	0	Aprecierea vietii in general (1-1	None	11 - 99	8	Right	Ordinal
18	a4	Numeric	1	0	Cat de fericit sunteti?	{1, Foarte fericit}...	5 - 8, 0	8	Right	Ordinal
19	a5prie	Numeric	1	0	Cat de fericit e cel mai bun pri	{1, Foarte nefericit}...	5 - 7, 0	8	Right	Ordinal
20	a5colegi	Numeric	1	0	Cat de fericiti sunt colegii de s	{1, Foarte nefericit}...	5 - 7, 0	8	Right	Ordinal
21	a5membri	Numeric	1	0	Cat de fericiti sunt membrii gru	{1, Foarte nefericit}...	5 - 7, 0	8	Right	Ordinal
22	a5fratii	Numeric	1	0	Cat de fericit sunt fratii?	{1, Foarte nefericit}...	5 - 7, 0	8	Right	Ordinal
23	a5parint	Numeric	1	0	Cat de fericit sunt parintii?	{1, Foarte nefericit}...	5 - 7, 0	8	Right	Ordinal
24	a5vecini	Numeric	1	0	Cat de fericit sunt vecinii?	{1, Foarte nefericit}...	5 - 7, 0	8	Right	Ordinal
25	a5locali	Numeric	1	0	Cat de fericit sunt cei din local	{1, Foarte nefericit}...	5 - 7, 0	8	Right	Ordinal
26	a5tara	Numeric	1	0	Cat de fericiti sunt concetateni	{1, Foarte nefericit}...	5 - 7, 0	8	Right	Ordinal
27	a6a	Numeric	1	0	Consumul - probleme sanatate	{1, Da}...	3 - 9, 0	8	Right	Ordinal
28	a6b	Numeric	1	0	Consumul - incalcarea legilor	{1, Da}...	3 - 9, 0	8	Right	Ordinal
29	a6c	Numeric	1	0	Consumul - probleme familie	{1, Da}...	3 - 9, 0	8	Right	Ordinal
30	a6d	Numeric	1	0	Consumul - saracie	{1, Da}...	3 - 9, 0	8	Right	Ordinal
31	a6e	Numeric	1	0	Consumul - violenta	{1, Da}...	3 - 9, 0	8	Right	Ordinal
32	a6f	Numeric	1	0	Consumul - sinucidere	{1, Da}...	3 - 9, 0	8	Right	Ordinal
33	a6g	Numeric	1	0	Consumul - probleme profesio	{1, Da}...	3 - 9, 0	8	Right	Ordinal
34	a6h	Numeric	1	0	Consumul - SIDA	{1, Da}...	3 - 9, 0	8	Right	Ordinal
35	a7	Numeric	1	0	Ai vazut consumator de drog?	{1, Da}...	3 - 9, 0	8	Right	Ordinal
36	a8	Numeric	1	0	Ai intilnit cineva care presupui	{1, Da}...	3 - 9, 0	8	Right	Ordinal
37	a9	Numeric	1	0	Ai auzit despre consumator de	{1, Da}...	3 - 9, 0	8	Right	Ordinal
38	a10	Numeric	1	0	Ai colegi de clasa cons. de dr	{1, Da}...	3 - 9, 0	8	Right	Ordinal
39	a11	Numeric	2	0	Nr. colegi cons. de drog	None	0	8	Right	Ordinal
40	a12	Numeric	1	0	Ai grup de prieteni?	{1, Da}...	3 - 9, 0	8	Right	Ordinal
41	a13	Numeric	3	0	Nr. grup prieteni	None	0	8	Right	Ordinal
42	a14	Numeric	3	0	Nr. cons. drog din grup de prie	None	None	8	Right	Ordinal
43	a15	Numeric	3	0	Nr. cons. drog alte cunostinte	None	None	8	Right	Ordinal
44	a16	Numeric	3	0	Procent cons. drog din Oradea	None	99	8	Right	Ordinal
45	a17a	Numeric	1	0	Consum Marijuana?	{1, Des}...	4 - 9, 0	8	Right	Ordinal
46	a17b	Numeric	1	0	Consum Hasis?	{1, Des}...	4 - 9, 0	8	Right	Ordinal

2. Introduceti la Row(s) variabila originala – a1
3. Introduceti la Column (s) variabila noua – a1rec
4. Apasati **OK**



In output se scrie rezultatul analizei de asociere: tabelul de asociere (vezi mai jos) care confirma corectitudinea comenzii de recodificare.



Comanda Crosstabs este una mult mai puternica, prin care se poate testa statistic relatia dintre doua variabile categorice. Pentru a realiza tabele de asociere simple, alternativa este utilizarea comenzii Tables din Analyze.

Tema 3: Crearea de variabile calculate

Problema: cunoscându-se rata de frecventare a teatrului şi cinematografului şi faptul că preţul mediu al unui bilet de teatru este de 10 de lei iar al unui bilet de cinema de 15 lei, să se estimeze cheltuielile anuale pentru teatru şi cinema ale fiecărui adolescent din baza de date.

Soluție

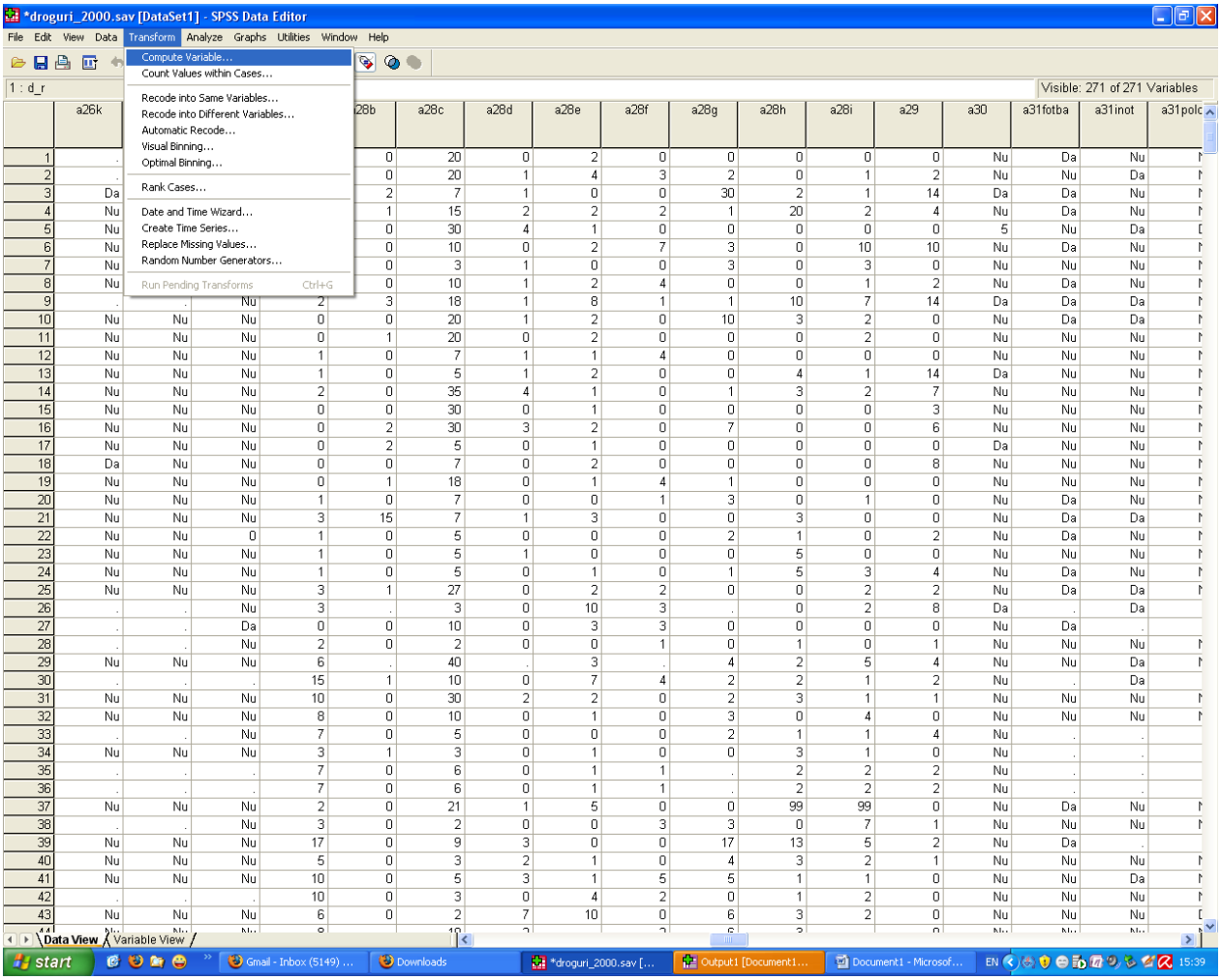
Variabilele utilizate sunt:

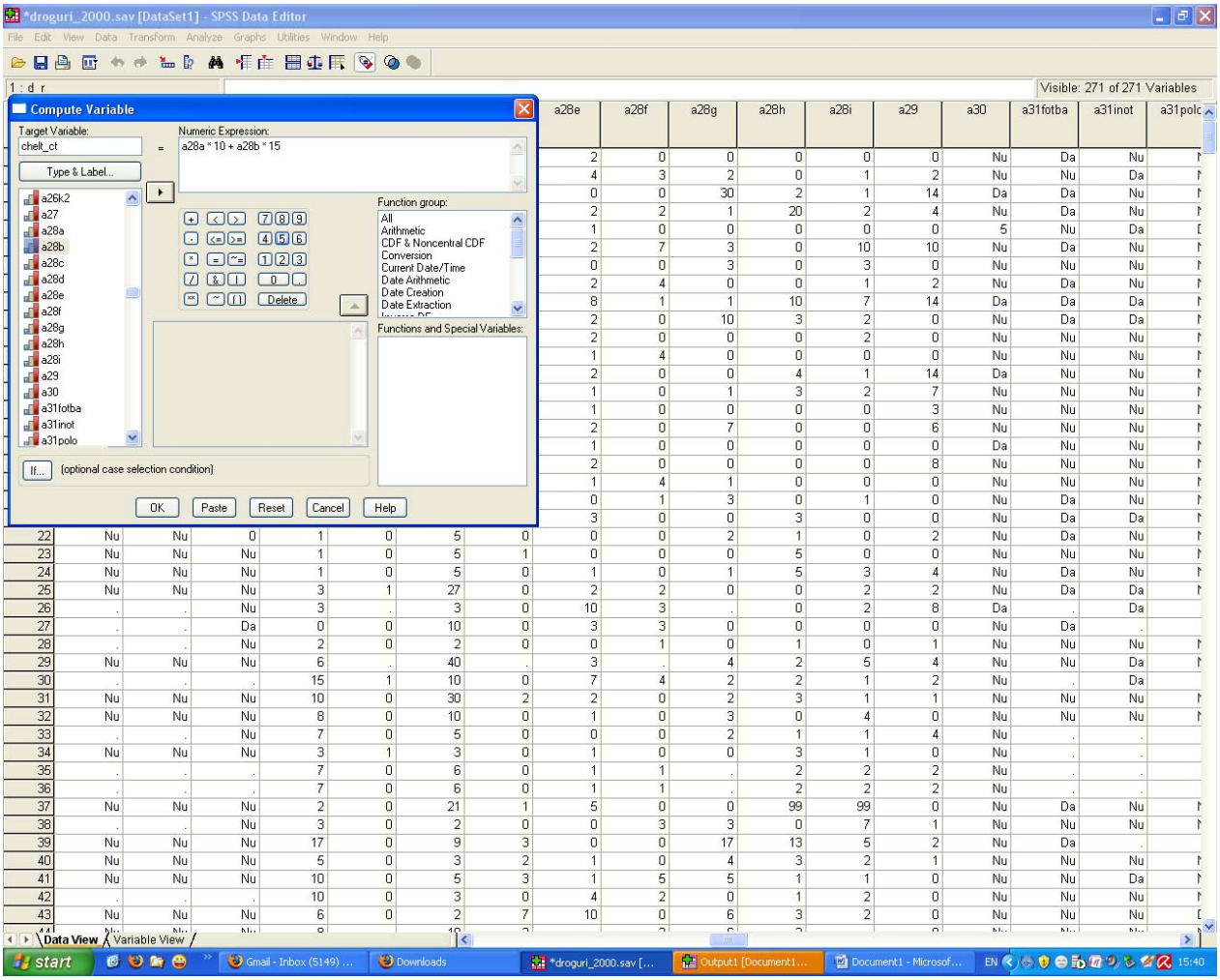
- a28a: frecvența teatru în ultimul an
- a28b: frecvența cinema în ultima lună

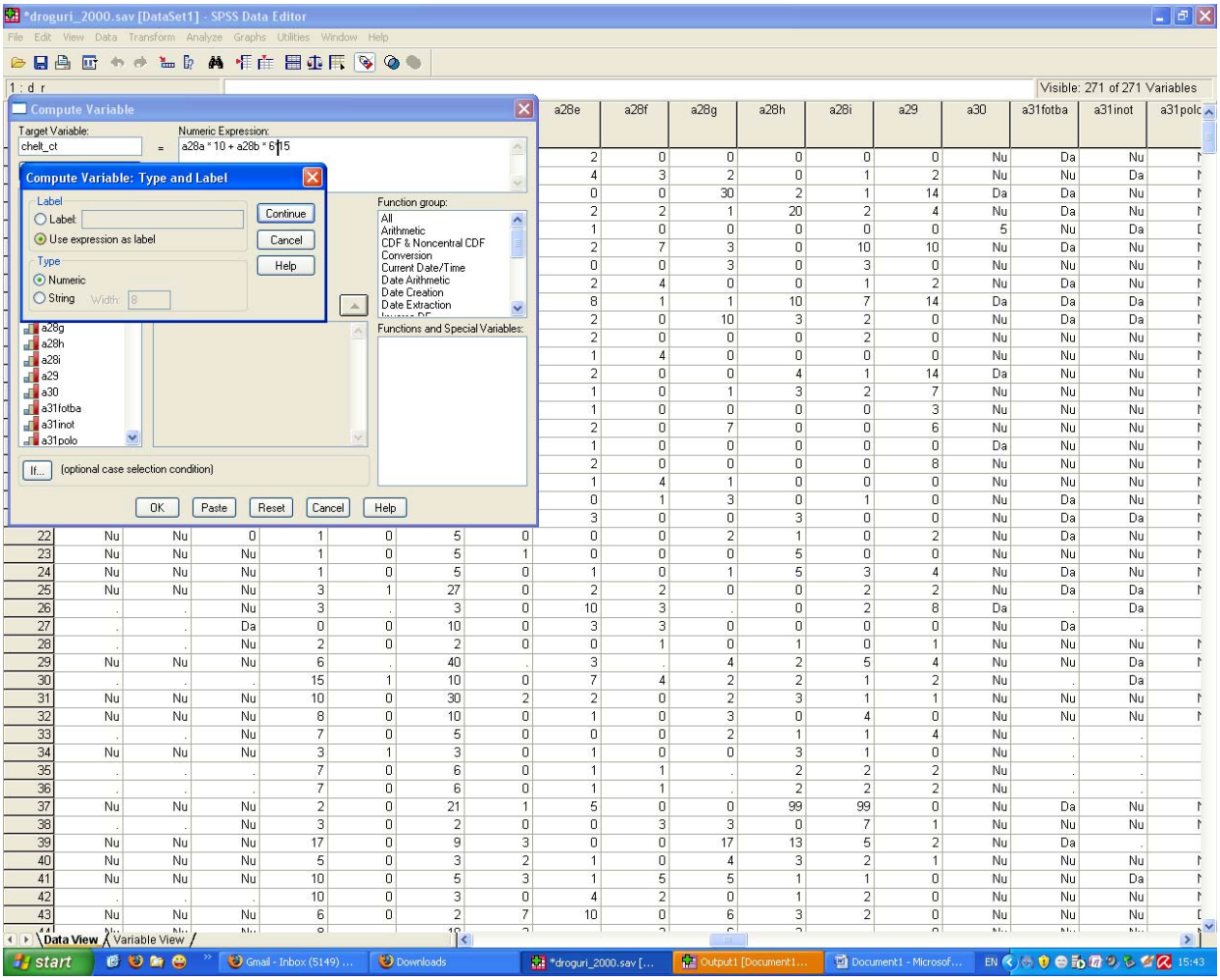
Suma de bani cheltuită pe an pentru cele două forme de petrecere a timpului liber se poate afla astfel= a28a*10+a28b*6*15

Pentru a genera o nouă variabilă în SPSS care să conțină această informație folosim opțiunea Compute din meniul Transform.

1. Alegem opțiunea Compute din meniul Transform
2. Completați în fereastra de dialog următoarele:
    a. Numele noii variabile (chelt_ct) la *Target variable*
    b. Expresia numerică a28a*10+a28b*6*15 la *Numeric expression* (înmulțirea cu 6 se face pentru a extinde calculul la un an întreg)
3. Apăsați butonul **Type and label** şi alegeți opțiunea **Use expression as label** (aceasta vă va ajuta să înțelegeți mai târziu semnificația noii variabile)
4. Apăsați **Continue**, apoi **OK**




*droguri_2000.sav [DataSet1] - SPSS Data Editor
File Edit View Data Transform Analyze Graphs Utilities Window Help
1 : d r
Visible: 271 of 271 Variables
Compute Variable
Target Variable:
chelt_ct
Type & Label...
Numeric Expression:
a28a * 10 + a28b * 15
Function group:
All
Arithmetic
CDF & Noncentral CDF
Conversion
Current Date/Time
Date Arithmetic
Date Creation
Date Extraction
Functions and Special Variables:
a26k2
a27
a28a
a28b
a28c
a28d
a28e
a28f
a28g
a28h
a28i
a29
a30
a31fotba
a31inot
a31polo
If... (optional case selection condition)
OK
Paste
Reset
Cancel
Help
Delete
a28e a28f a28g a28h a28i a29 a30 a31fotba a31inot a31polc
Data View
Variable View
start
Gmail - Inbox (5149) ...
Downloads
*droguri_2000.sav [...
Output1 [Document1...
Document1 - Microsof...
EN
15:40



La capătul bazei de date v-a răsărit noua variabilă: chelt_tc

La folosirea acestei comenzi trebuie avută în vedere prezența unor valori aberante ori a unor coduri speciale, precum 99-nu răspund, care trebuiesc eliminate pentru a nu produce rezultate bizare în noua variabilă.

Tema 4: Numărarea ocurenței unor evenimente/valori dintr-un set de variabile într-o nouă variabilă

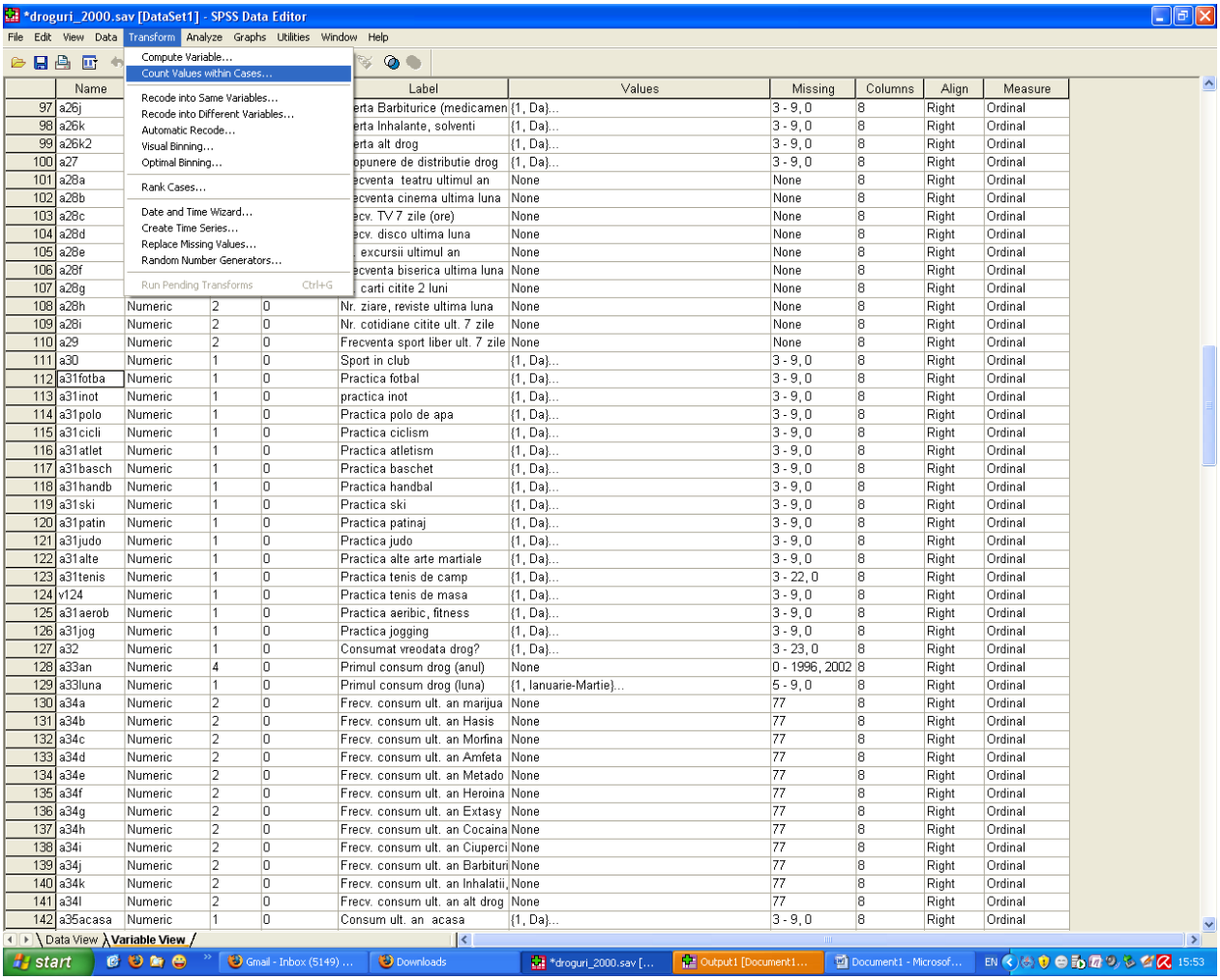
Problema: câte sporturi practică fiecare dintre elevii din baza de date

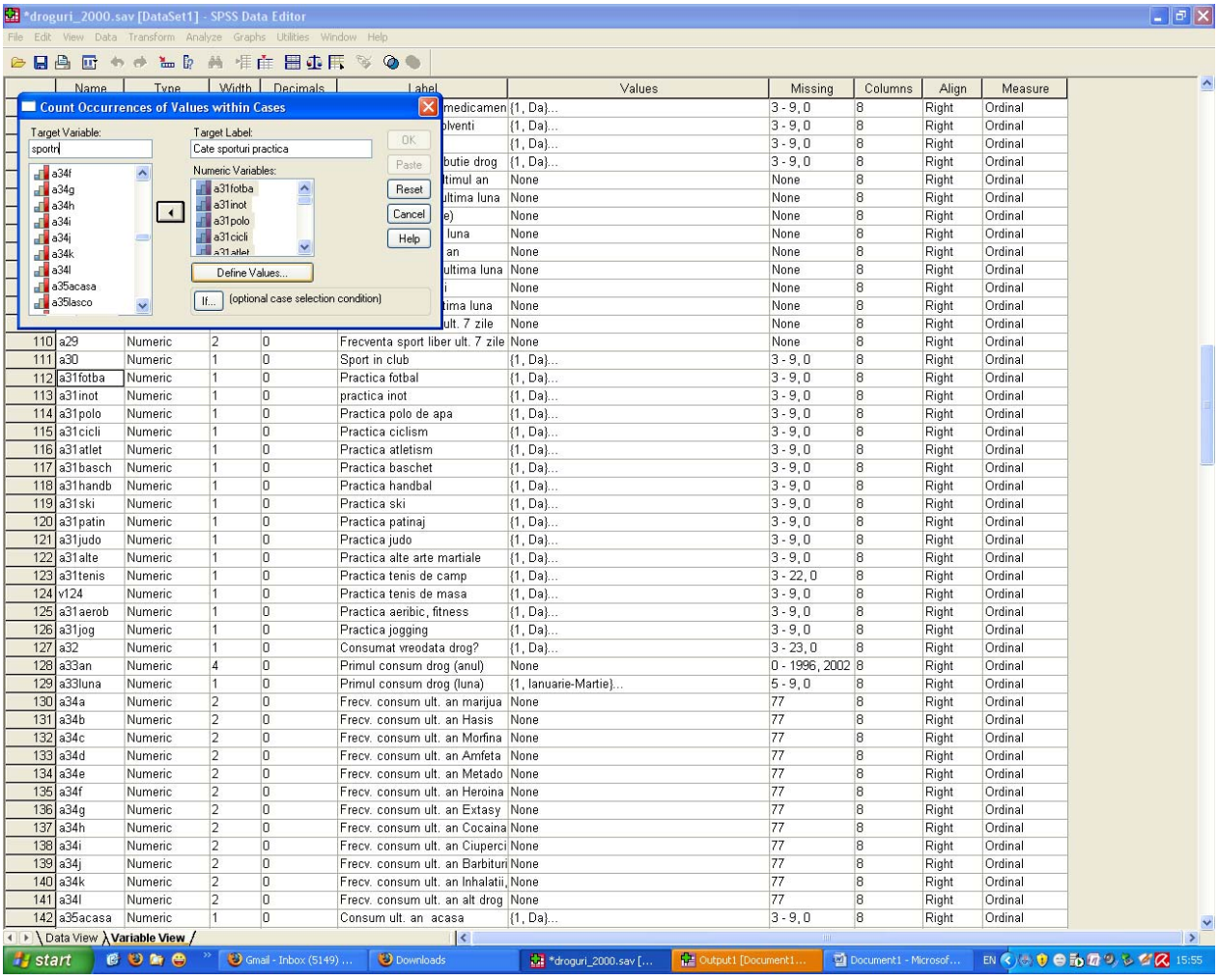
Soluție

Elevii au fost întrebați dacă practică 15 sporturi, care pot fi găsite în baza de date de la a31fotba până la a31jog. Pentru fiecare sport, se putea răspunde da (cod 1), adică practică sportul respectiv sau nu (cod 2). A afla câte sporturi practică fiecare înseamnă a număra de câte ori a răspuns da la cele 15 întrebări. Cu alte cuvinte, soluția constă în a afla de câte ori avem codul 1 în întrebările de la a31fotba până la

a31jog. Pentru a genera o nouă variabilă care va cuprinde această informație, vom folosi opțiunea Count.

1. Din meniul *Transform* alegeți opțiunea *Count values within cases*
2. În fereastra de dialog introduceți:
    a. Numele noii variabile (*sportn*)
    b. Listade variabile în care se face numărarea (a31fotba până la a31jog)
    c. Eticheta noii variabile
    d. Apăsați butonul **Define values**

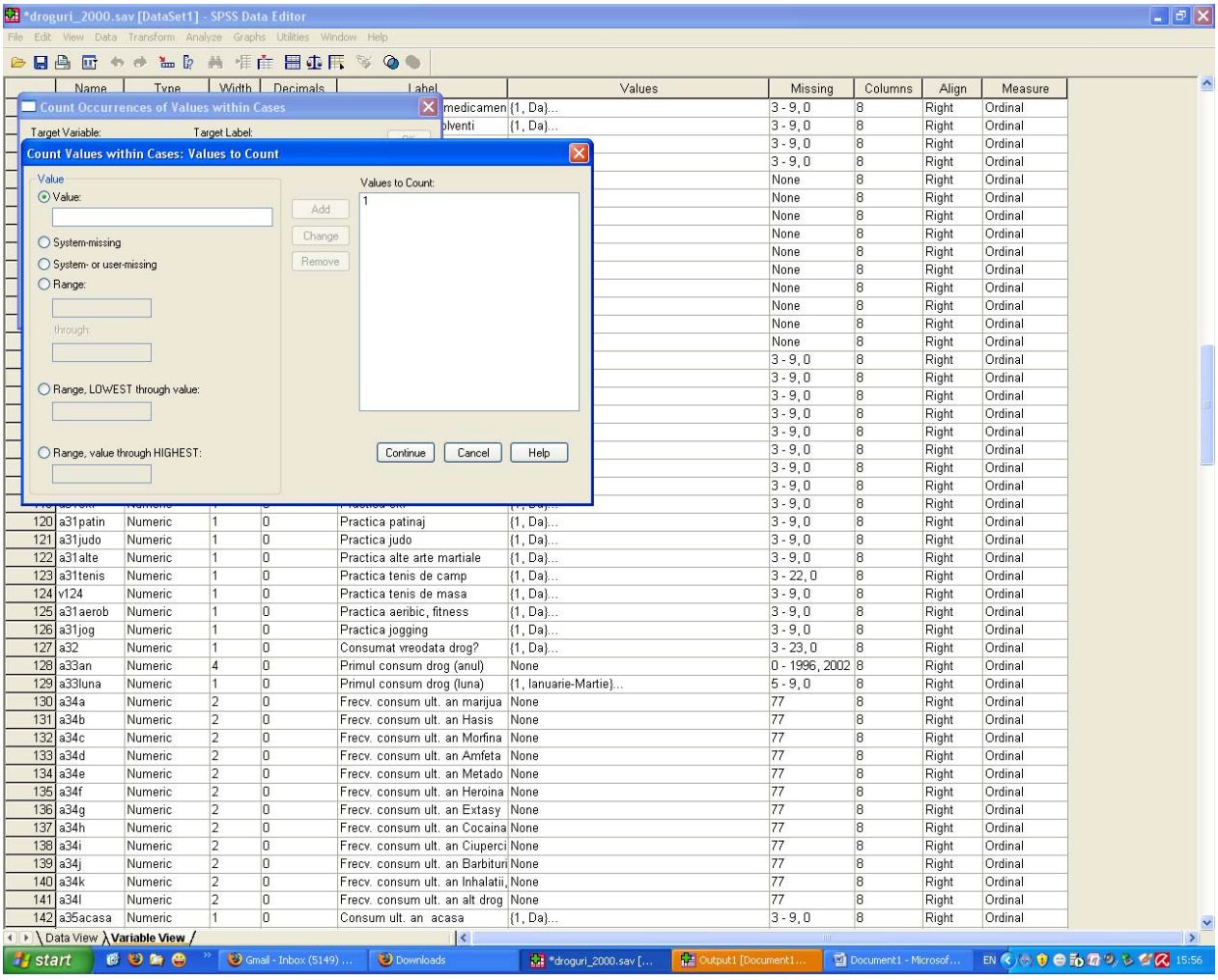




	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
110	a29	Numeric	2	0	Frecventa sport liber ult. 7 zile	None	None	8	Right	Ordinal
111	a30	Numeric	1	0	Sport in club	{1, Da}...	3 - 9, 0	8	Right	Ordinal
112	a31fotba	Numeric	1	0	Practica fotbal	{1, Da}...	3 - 9, 0	8	Right	Ordinal
113	a31inot	Numeric	1	0	practica inot	{1, Da}...	3 - 9, 0	8	Right	Ordinal
114	a31polo	Numeric	1	0	Practica polo de apa	{1, Da}...	3 - 9, 0	8	Right	Ordinal
115	a31cicli	Numeric	1	0	Practica ciclism	{1, Da}...	3 - 9, 0	8	Right	Ordinal
116	a31atlet	Numeric	1	0	Practica atletism	{1, Da}...	3 - 9, 0	8	Right	Ordinal
117	a31basch	Numeric	1	0	Practica baschet	{1, Da}...	3 - 9, 0	8	Right	Ordinal
118	a31handb	Numeric	1	0	Practica handbal	{1, Da}...	3 - 9, 0	8	Right	Ordinal
119	a31ski	Numeric	1	0	Practica ski	{1, Da}...	3 - 9, 0	8	Right	Ordinal
120	a31patin	Numeric	1	0	Practica patinaj	{1, Da}...	3 - 9, 0	8	Right	Ordinal
121	a31judo	Numeric	1	0	Practica judo	{1, Da}...	3 - 9, 0	8	Right	Ordinal
122	a31alte	Numeric	1	0	Practica alte arte martiale	{1, Da}...	3 - 9, 0	8	Right	Ordinal
123	a31tenis	Numeric	1	0	Practica tenis de camp	{1, Da}...	3 - 22, 0	8	Right	Ordinal
124	v124	Numeric	1	0	Practica tenis de masa	{1, Da}...	3 - 9, 0	8	Right	Ordinal
125	a31aerob	Numeric	1	0	Practica aeribic, fitness	{1, Da}...	3 - 9, 0	8	Right	Ordinal
126	a31jog	Numeric	1	0	Practica jogging	{1, Da}...	3 - 9, 0	8	Right	Ordinal
127	a32	Numeric	1	0	Consumat vreodata drog?	{1, Da}...	3 - 23, 0	8	Right	Ordinal
128	a33an	Numeric	4	0	Primul consum drog (anul)	None	0 - 1996, 2002	8	Right	Ordinal
129	a33luna	Numeric	1	0	Primul consum drog (luna)	{1, Ianuarie-Martie}...	5 - 9, 0	8	Right	Ordinal
130	a34a	Numeric	2	0	Frecv. consum ult. an marijua	None	77	8	Right	Ordinal
131	a34b	Numeric	2	0	Frecv. consum ult. an Hasis	None	77	8	Right	Ordinal
132	a34c	Numeric	2	0	Frecv. consum ult. an Morfina	None	77	8	Right	Ordinal
133	a34d	Numeric	2	0	Frecv. consum ult. an Amfeta	None	77	8	Right	Ordinal
134	a34e	Numeric	2	0	Frecv. consum ult. an Metado	None	77	8	Right	Ordinal
135	a34f	Numeric	2	0	Frecv. consum ult. an Heroina	None	77	8	Right	Ordinal
136	a34g	Numeric	2	0	Frecv. consum ult. an Extasy	None	77	8	Right	Ordinal
137	a34h	Numeric	2	0	Frecv. consum ult. an Cocaina	None	77	8	Right	Ordinal
138	a34i	Numeric	2	0	Frecv. consum ult. an Ciuperci	None	77	8	Right	Ordinal
139	a34j	Numeric	2	0	Frecv. consum ult. an Barbituri	None	77	8	Right	Ordinal
140	a34k	Numeric	2	0	Frecv. consum ult. an Inhalatii	None	77	8	Right	Ordinal
141	a34l	Numeric	2	0	Frecv. consum ult. an alt drog	None	77	8	Right	Ordinal
142	a35acasa	Numeric	1	0	Consum ult. an acasa	{1, Da}...	3 - 9, 0	8	Right	Ordinal

În fereastra de dialog *Values to Count*, introduceți valoare 1, care va fi numărată.

Apăsați butonul **Continue** apoi **OK** .



La capătul bazei de date a apărut noua variabilă.

Succes!

Distribuții de frecvență și grafice cu variabile categorice

Conținut:

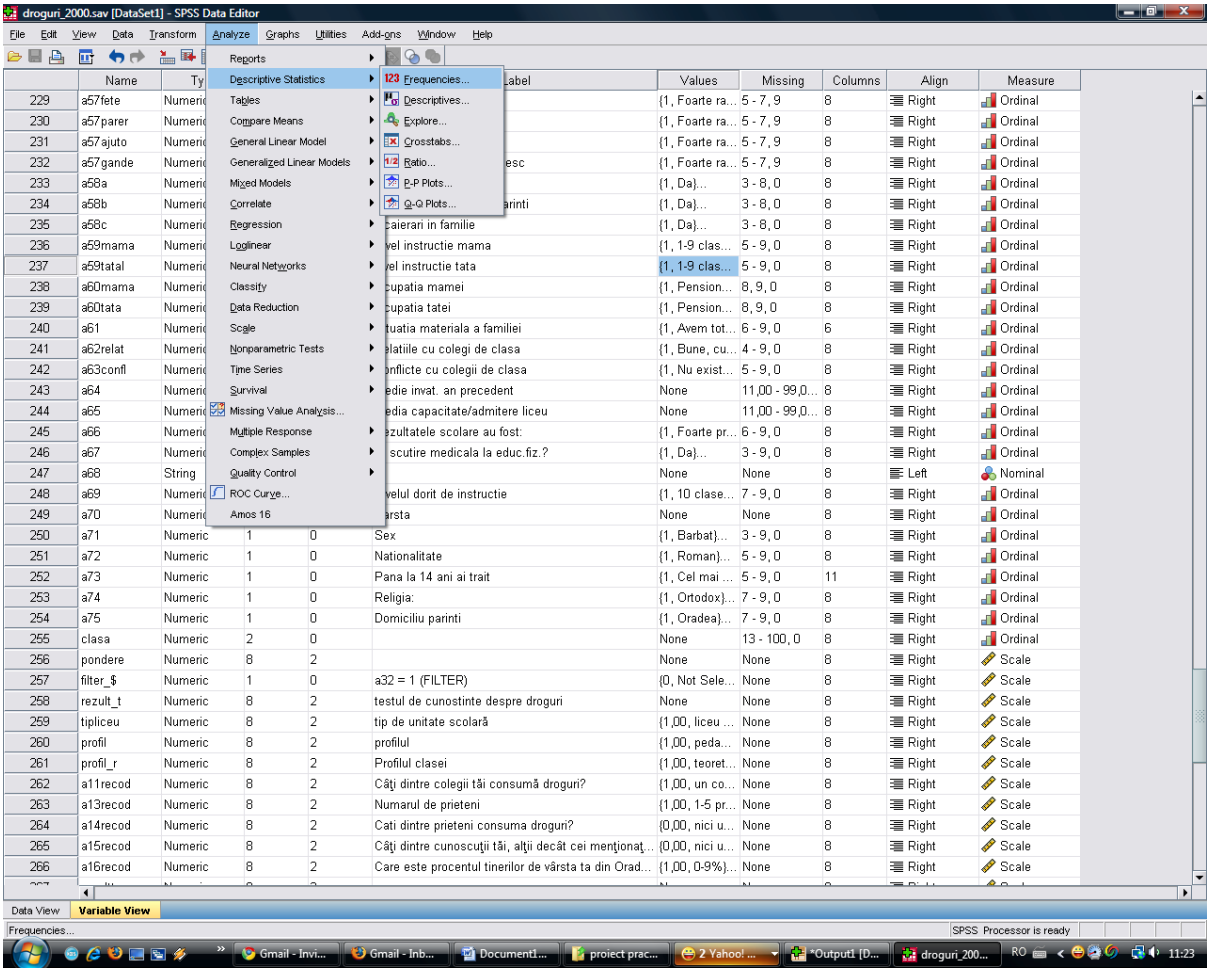
- **Distribuții de frecvență**
- **Grafice în SPSS**
- **Distribuții de frecvență cu Custom Tables**
- **Grafice în Excel 2007 - sinteză**

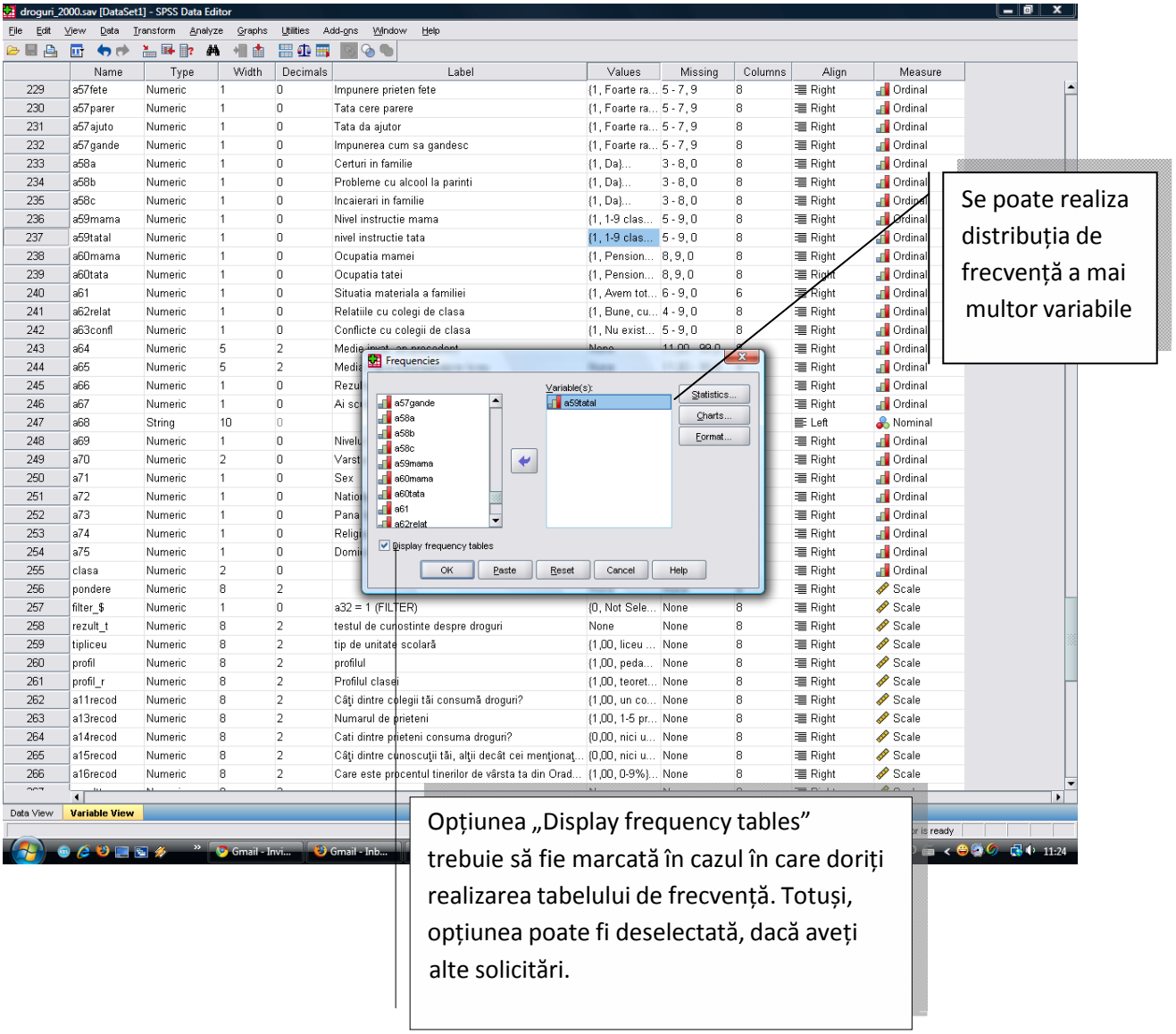
Problemă: descrieți elevii din baza de date droguri2000.sav după nivelul de instrucție al tatălui.

Soluție

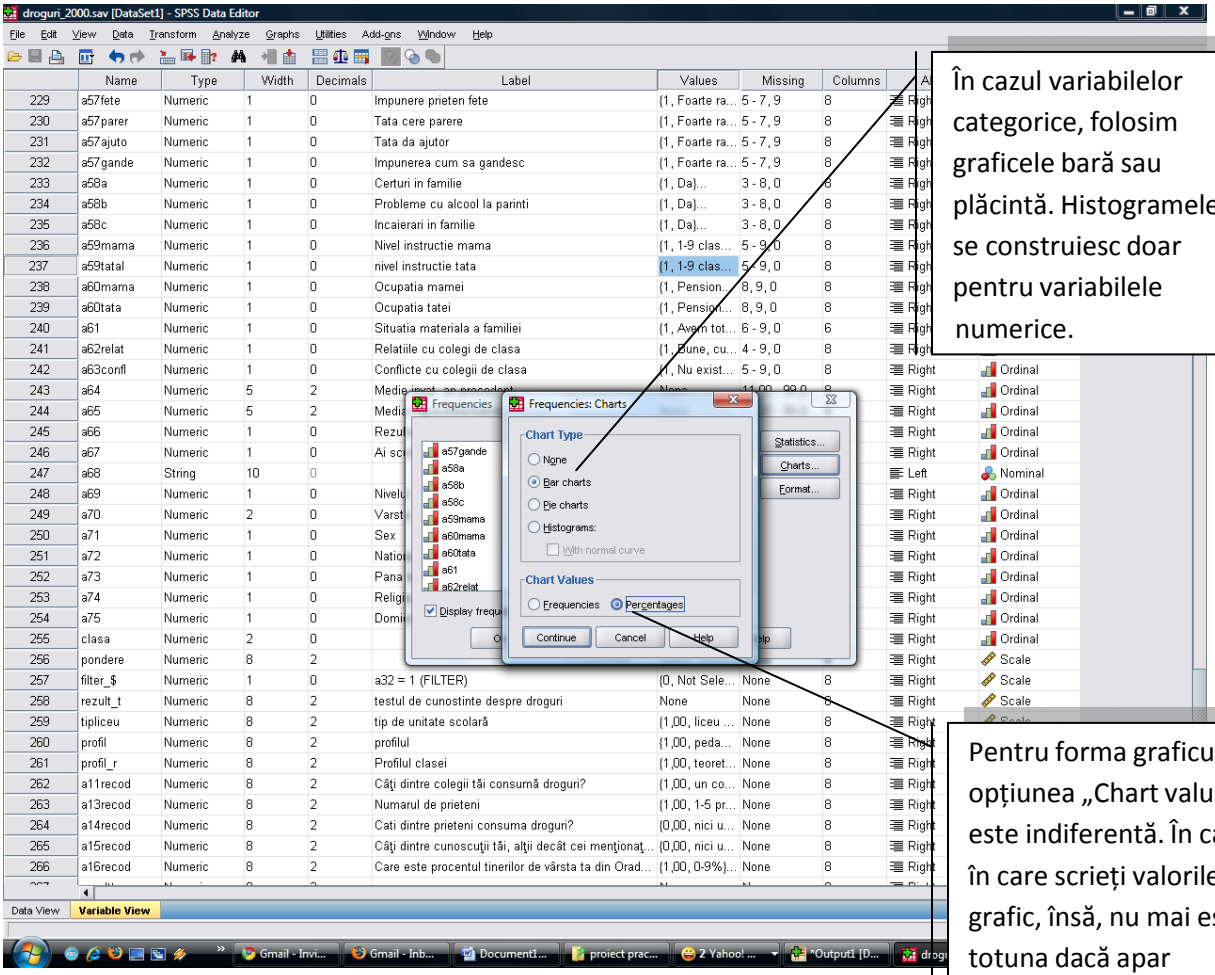
Descrierea variabilelor categorice se face cel mai bine cu comanda *Frequencies*. Aceasta permite realizarea de:

- Tabele de frecvență
- Grafice bară și plăcintă
- Statistici descriptive parametrice


droguri_2000.sav [DataSet1] - SPSS Data Editor
File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help
Reports
Descriptive Statistics
Tables
Compare Means
General Linear Model
Generalized Linear Models
Mixed Models
Correlate
Regression
Loglinear
Neural Networks
Classify
Data Reduction
Scale
Nonparametric Tests
Time Series
Survival
Missing Value Analysis...
Multiple Response
Complex Samples
Quality Control
ROC Curve...
Amos 16
Frequencies...
Descriptives...
Explore...
Crosstabs...
Ratio...
P-P Plots...
Q-Q Plots...
Name Ty Label Values Missing Columns Align Measure
229 a57fete Numeric {1, Foarte ra... 5 - 7, 9 8 Right Ordinal
230 a57parer Numeric {1, Foarte ra... 5 - 7, 9 8 Right Ordinal
231 a57ajuto Numeric {1, Foarte ra... 5 - 7, 9 8 Right Ordinal
232 a57gande Numeric {1, Foarte ra... 5 - 7, 9 8 Right Ordinal
233 a58a Numeric {1, Da}... 3 - 8, 0 8 Right Ordinal
234 a58b Numeric {1, Da}... 3 - 8, 0 8 Right Ordinal
235 a58c Numeric {1, Da}... 3 - 8, 0 8 Right Ordinal
236 a59mama Numeric {1, 1-9 clas... 5 - 9, 0 8 Right Ordinal
237 a59tatal Numeric {1, 1-9 clas... 5 - 9, 0 8 Right Ordinal
238 a60mama Numeric {1, Pension... 8, 9, 0 8 Right Ordinal
239 a60tata Numeric {1, Pension... 8, 9, 0 8 Right Ordinal
240 a61 Numeric {1, Avem tot... 6 - 9, 0 6 Right Ordinal
241 a62relat Numeric {1, Bune, cu... 4 - 9, 0 8 Right Ordinal
242 a63confl Numeric {1, Nu exist... 5 - 9, 0 8 Right Ordinal
243 a64 Numeric None 11,00 - 99,0... 8 Right Ordinal
244 a65 Numeric None 11,00 - 99,0... 8 Right Ordinal
245 a66 Numeric {1, Foarte pr... 6 - 9, 0 8 Right Ordinal
246 a67 Numeric {1, Da}... 3 - 9, 0 8 Right Ordinal
247 a68 String None None 8 Left Nominal
248 a69 Numeric {1, 10 clase... 7 - 9, 0 8 Right Ordinal
249 a70 Numeric None None 8 Right Ordinal
250 a71 Numeric 1 0 Sex {1, Barbat}... 3 - 9, 0 8 Right Ordinal
251 a72 Numeric 1 0 Nationalitate {1, Roman}... 5 - 9, 0 8 Right Ordinal
252 a73 Numeric 1 0 Pana la 14 ani ai trait {1, Cel mai ... 5 - 9, 0 11 Right Ordinal
253 a74 Numeric 1 0 Religia: {1, Ortodox}... 7 - 9, 0 8 Right Ordinal
254 a75 Numeric 1 0 Domiciliu parinti {1, Oradea}... 7 - 9, 0 8 Right Ordinal
255 clasa Numeric 2 0 None 13 - 100, 0 8 Right Ordinal
256 pondere Numeric 8 2 None None 8 Right Scale
257 filter_$ Numeric 1 0 a32 = 1 (FILTER) {0, Not Sele... None 8 Right Scale
258 rezult_t Numeric 8 2 testul de cunostinte despre droguri None None 8 Right Scale
259 tipliceu Numeric 8 2 tip de unitate scolară {1,00, liceu ... None 8 Right Scale
260 profil Numeric 8 2 profilul {1,00, peda... None 8 Right Scale
261 profil_r Numeric 8 2 Profilul clasei {1,00, teoret... None 8 Right Scale
262 a11recod Numeric 8 2 Câţi dintre colegii tăi consumă droguri? {1,00, un co... None 8 Right Scale
263 a13recod Numeric 8 2 Numarul de prieteni {1,00, 1-5 pr... None 8 Right Scale
264 a14recod Numeric 8 2 Cati dintre prieteni consuma droguri? {0,00, nici u... None 8 Right Scale
265 a15recod Numeric 8 2 Câţi dintre cunoscuţii tăi, alţii decât cei menţionaţ... {0,00, nici u... None 8 Right Scale
266 a16recod Numeric 8 2 Care este procentul tinerilor de vârsta ta din Orad... {1,00, 0-9%}... None 8 Right Scale
Data View Variable View
Frequencies...
SPSS Processor is ready
Gmail - Invi... Gmail - Inb... Document1... proiect prac... 2 Yahoo! ... *Output1 [D... droguri_200... RO 11:23



Odată cu tabelul de frecvență se poate solicita și realizarea unor grafice. Pentru aceasta se apasă butonul Charts ceea ce duce la deschiderea unei ferestre de dialog. În cazul variabilelor categorice se pot realiza grafice bară (*Bar charts*) și plăcintă (*Pie charts*).



Apăsăm Continue apoi OK iar rezultatul este scris în output:

Rezultate

Număr de cazuri valide și lipsă

Statistics

nivel instructie tata

N	Valid	1081
	Missing	49

nivel instructie tata

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1-9 clase	70	6,2	6,5	6,5
	10 cl. /scoala. prof.-la	249	22,0	23,0	29,5
	Liceu, bac., postliceal	426	37,7	39,4	69,0
	Studii superioare	335	29,7	31,0	100,0
	Total	1081	95,6	100,0	
Missing	0	2	,2		
	6	1	,1		
	7	1	,1		
	System	45	4,0		
	Total	49	4,4		
Total		1130	100,0		

Frecvențe. Avem 335 de elevi cu tatăl deținător de diplomă de studii superioare.

Numărul cazurilor valide este 1081 iar cel al cazurilor lipsă de 49

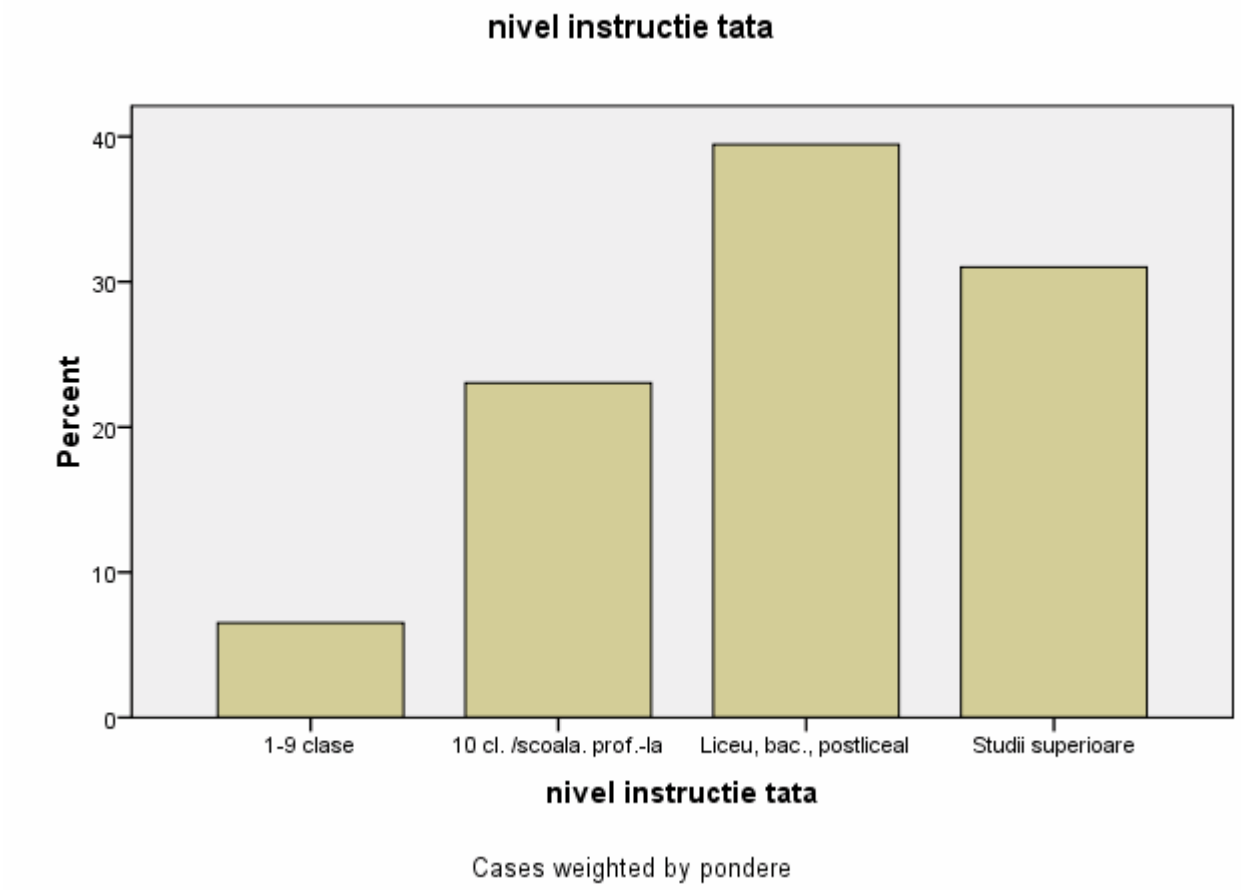
Procente din total:cei al căror tată are doar 1-9 clase reprezintă 100*70/1130 adică 6,2% din totalul cazurilor

Procente din totalul cazurilor valide (totalul după eliminarea cazurilor lipsă). Procentele din această coloană sunt, evident, mai mari decât cele din coloana anterioară.

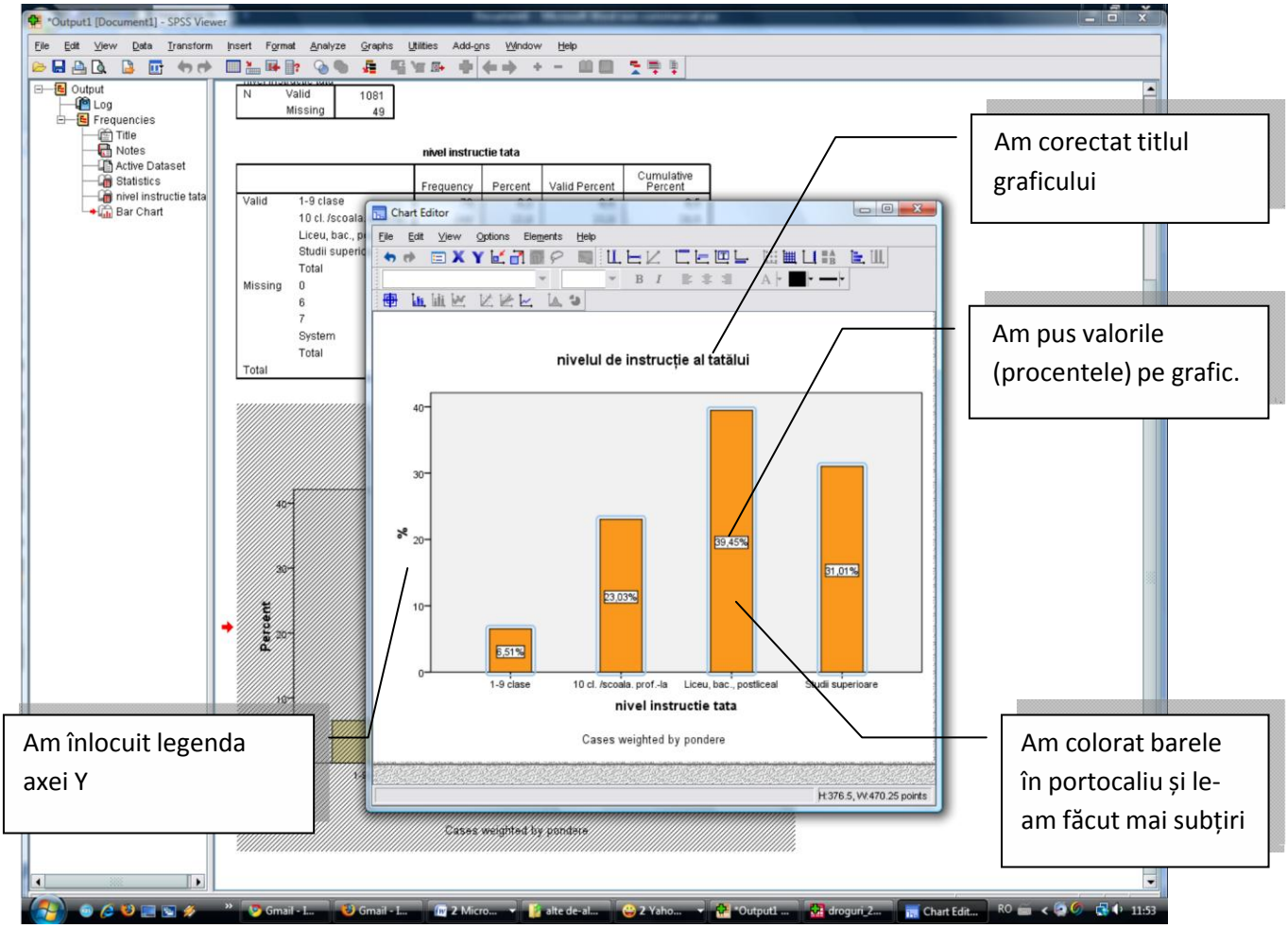
Procente cumulative. În cazul în care variabila este cel puțin ordinală, lectura procentelor cumulative poate fi folositoare pentru identificarea medianei.

Frecvența validă a categoriei se adună la frecvența cumulativă a valorii anterioare.

În cazul variabilelor care pot fi ordonate, ne permite identificarea medianei după regula: mediana este valoarea la care procentul cumulat depășește 50%.



Graficul cu bare de mai sus poate fi editat după ce faceți dublu-click pe graphic, comandă care duce la deschiderea editorului de grafice.

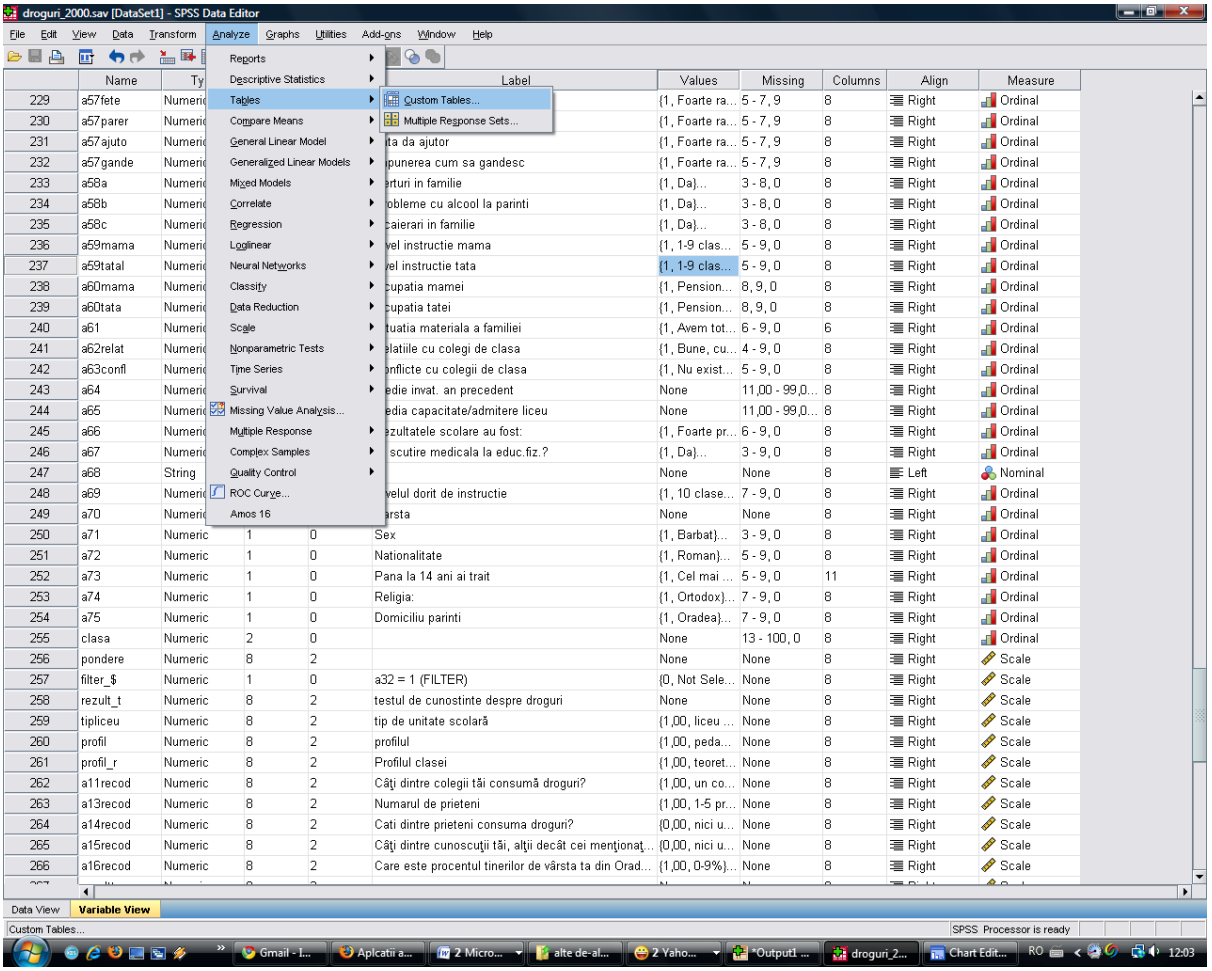


Totuși, pentru rapoarte arătoase, recomand realizarea tabelelor de frecvență cu comanda Custom Tables și construirea graficelor cu programul Excel (cel mai bine din pachetul Office 2007).

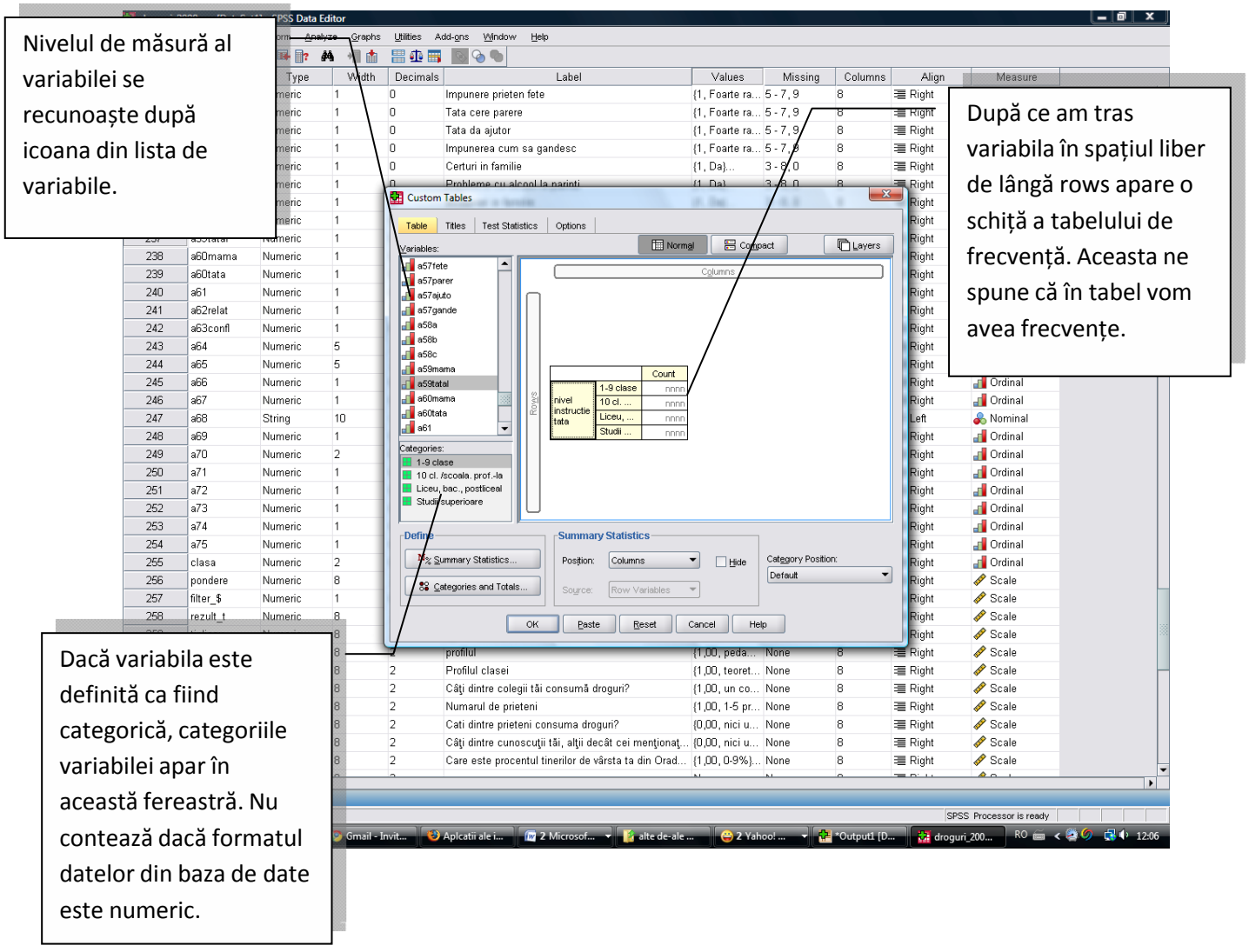
Tabele de frecven ă și grafice frumoase

Pentru a realiza tabele de frecvență cu comanda Tables, trebuie instalat modulul Tables din SPSS. Acesta nu este inclus în modulul de bază.

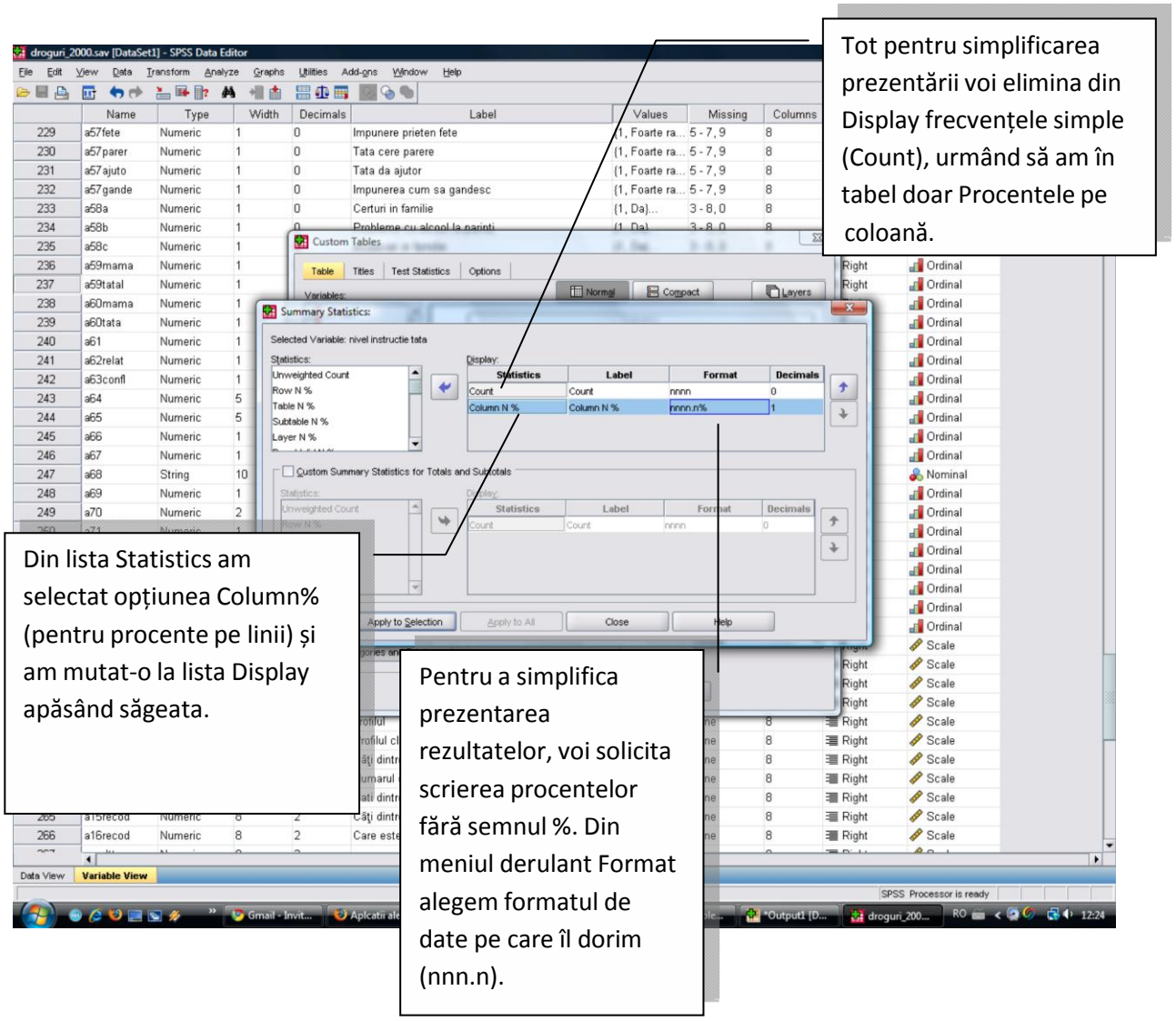
Opțiunile din submeniul *Tables* s-au schimbat de-a lungul timpului în diferitele versiuni de SPSS. Este de preferat comanda *Custom tables* deoarece are opțiunile cele mai flexibile.



1. Selectăm variabila a59tatal din lista de variabile având grijă ca aceasta să fie definită ca având măsurare nominală. Nivelul de măsură se recunoaște după icoana variabilei: dacă variabila este categorică va apare un mic grafic cu bare colorate. Dacă variabila este numerică, icoana va fi în forma unei rigle (care semnifică, probabil, măsurare metrică). În cazul în care variabila este definită ca fiind nominală, în secțiunea Categories vor apare valorile variabilei și etichetele acestora.
2. Mutăm, cu drag and drop (trage și aruncă, cu mouse-ul) variabila în dreptul spațiului în care scrie Rows. Va apare o schiță a tabelului de frecvență.



Dacă dorim schimbarea frecvențelor cu procente, vom fi nevoiți să apăsăm butonul Summary Statistics aflat în partea din stânga-jos a ferestrei de dialog Custom Tables.



Pentru a executa comanda apăsați **Apply to selection** apoi **OK**.

Rezultatul cuprinde doar procentele valide. Tabelul de mai jos poate fi folosit pentru a realiza cu ușurință grafice de bună calitate în Excel 2007 după algoritmul de mai jos:

- Se copiază tabelul într-o foaie de calcul excel cu **copy-paste**
- Se rotunjesc numerele (numărul de zecimale este prea mare pentru un grafic aspectuos)
- Se selectează zona de date
- Se selectează opțiunea **insert chart** și se stabilește tipul și formatul de grafic.
- Se editează zonele editabile (titlul, etichete etc.)

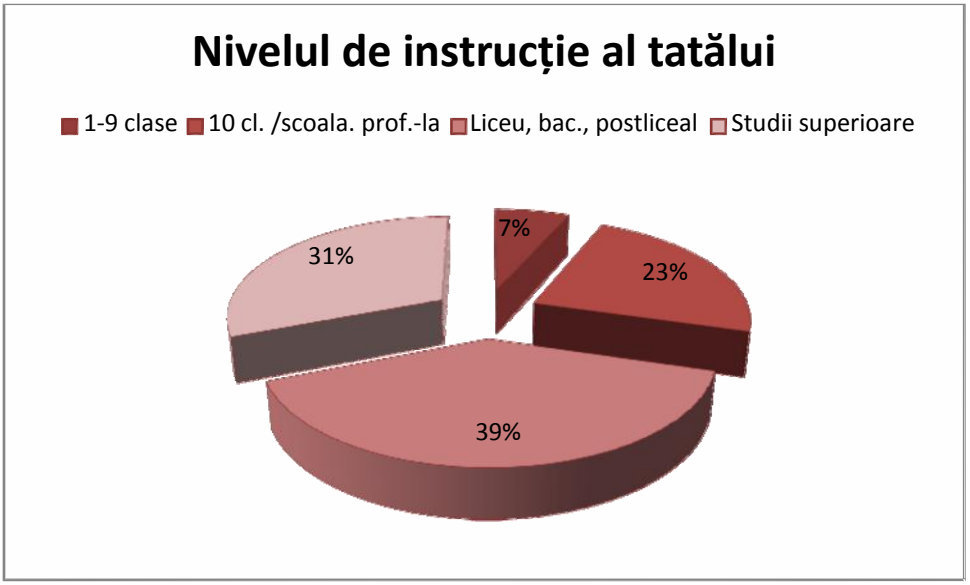
Table 1

		Column N %
nivel instructie tata	1-9 clase	6,5
	10 cl. /scoala. prof.-la	23,0
	Liceu, bac., postliceal	39,4

Table 1

		Column N %
nivel instructie tata	1-9 clase	6,5
	10 cl. /scoala. prof.-la	23,0
	Liceu, bac., postliceal	39,4
	Studii superioare	31,0

Iată un grafic făcut, simplu, în Excel 2007.



Descrierea variabilelor numerice

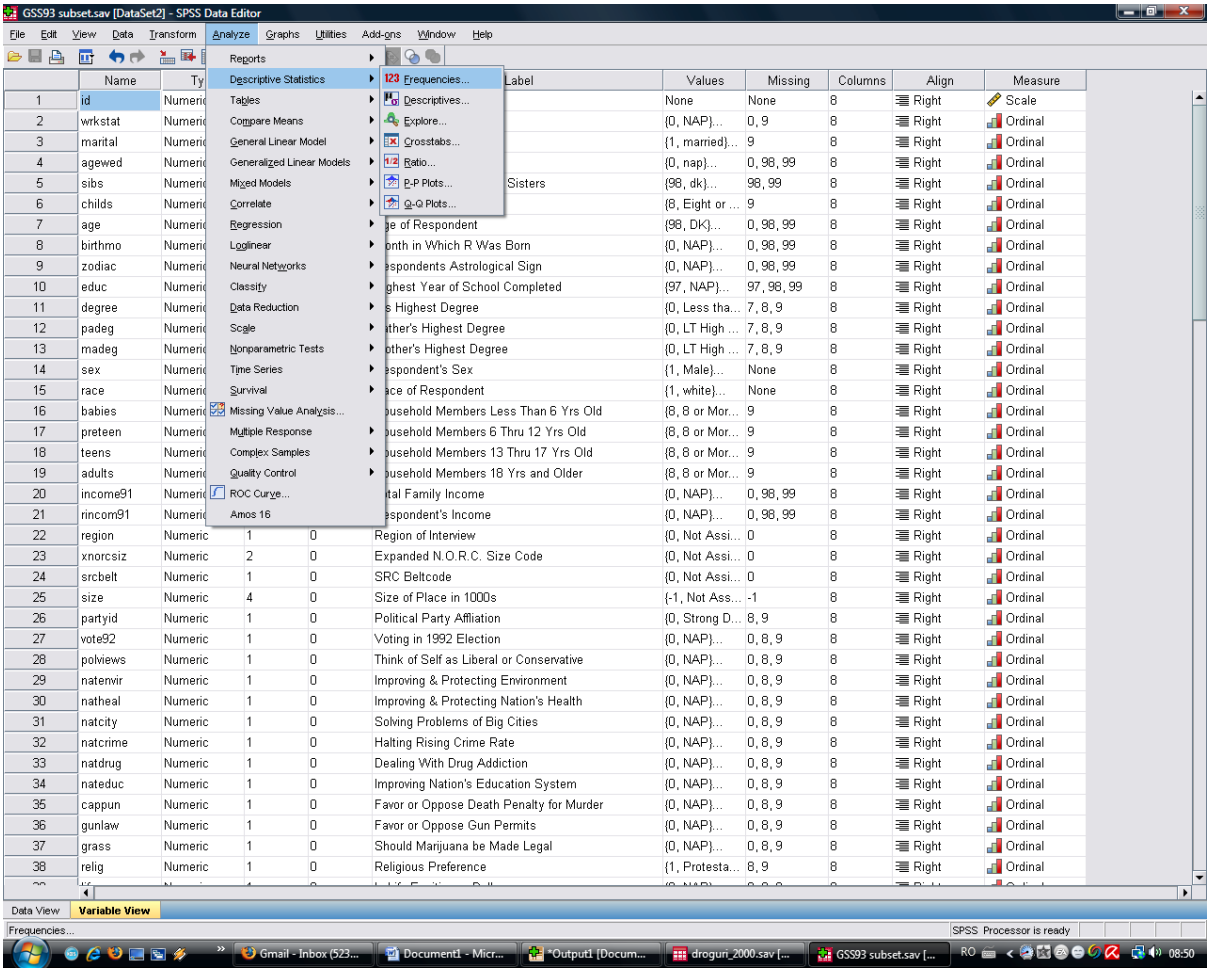
Pentru completarea cursului, recomand să citiți capitolul corespunzător din manualul de statistică al lui Pop și colaboratorii (2001) sau, măcar, prezentarea ppt de pe această pagină.

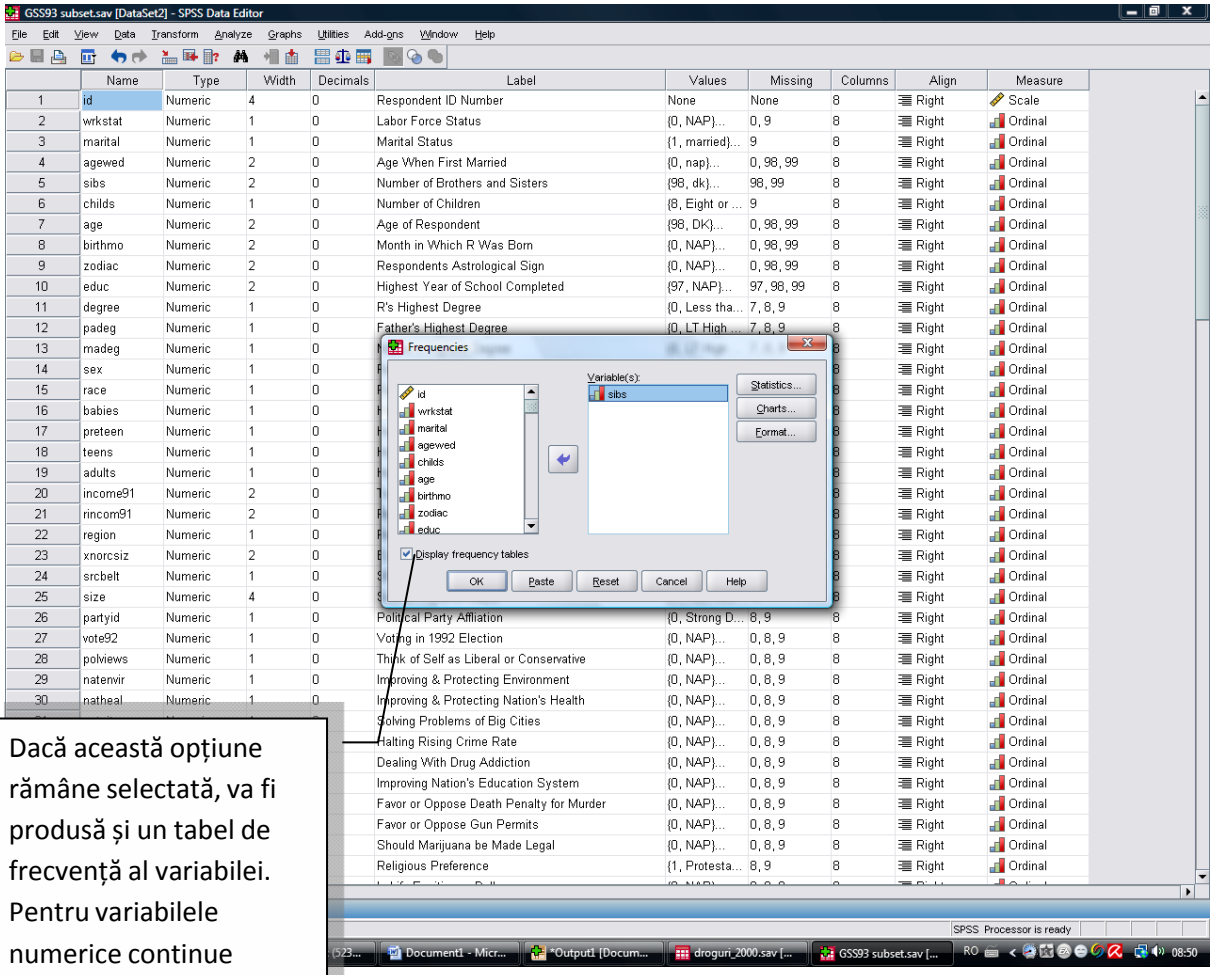
Problemă: să se descrie eșantionul american din *GSS93 subset* după numărul de frați al subiectului (variabila *sibs*)

Soluție

Variabila număr de frați este o variabilă numerică. ÎN cazul variabilelor numerice descrierile univariate sunt oarecum diferite de cele categorice.

SPSS permite mai multe proceduri de descriee parametrică univariată. În acest curs voi prezenta utilizarea comenzii Frequencies pentru acest scop și mai succint alte doouă comenzi: Descriptives și Explore. Toate aceste trei comenzi se găsesc în submeniul Descriptive Statistics.



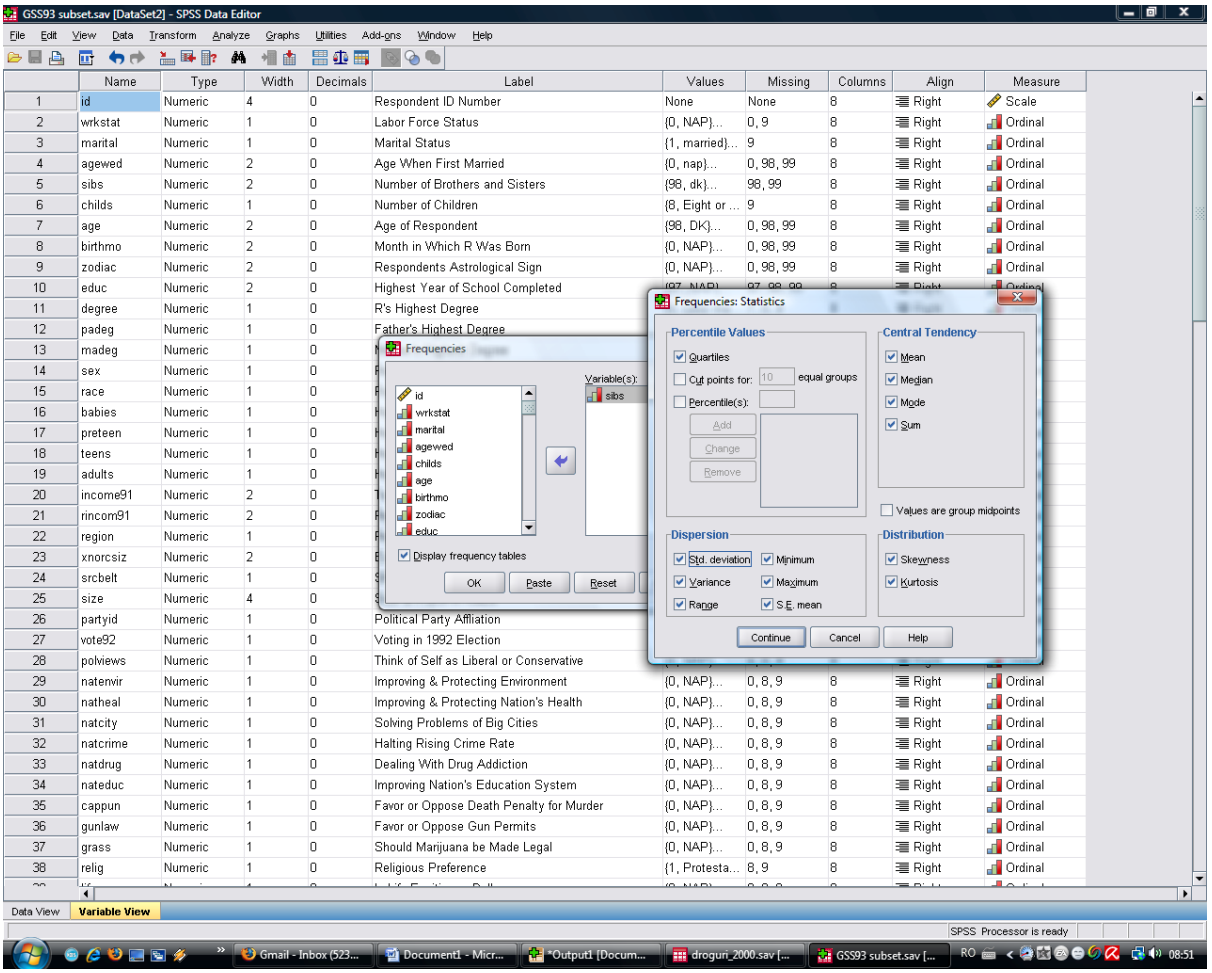


Dacă această opțiune rămâne selectată, va fi produsă și un tabel de frecvență al variabilei. Pentru variabilele numerice continue aceasta nu este o opțiune potrivită.

Calcularea indicilor statistici parametrici presupune activarea comenzii **Statistics**. Realizarea histogramei presupune activarea comenzii **Charts**.

După apăsarea butonului **Statistics** se deschide o fereastră de dialog în care trebuie marcate statisticile dorite. Acestea sunt gupate în cele patru categorii cunoscute:

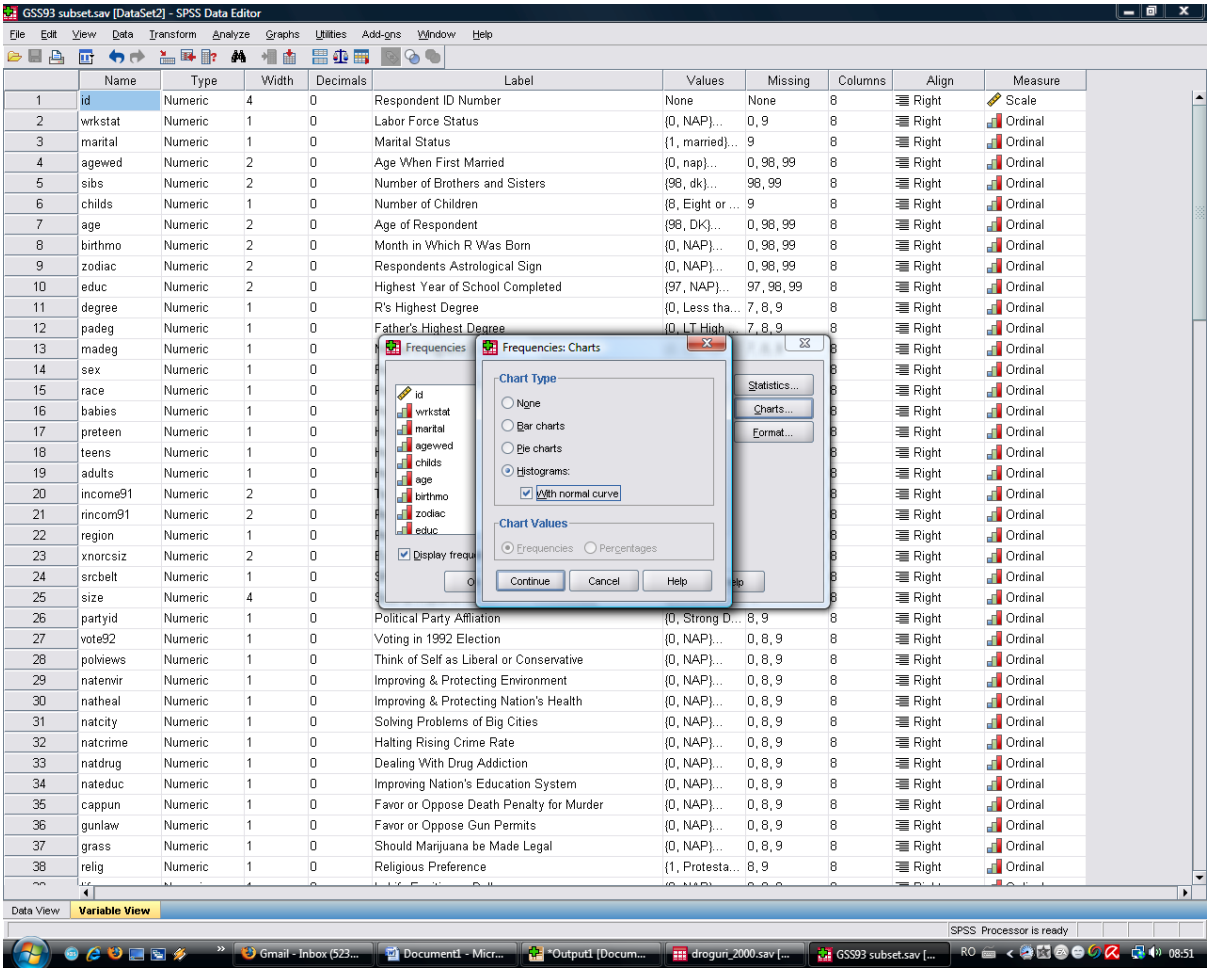
- Valori percentile
- Indicatori ai tendinței centrale
- Indicatori ai împrăștierii (dispersiei)
- Indicatori ai formei distribuției



După marcarea tuturor indicatorilor doriți apăsați **Continue.**

Din fereastra principală deschideți opțiunea **Charts**.

Selectați opțiunea Histogram și subopțiunea **With Normal Curve**. Aceasta vă va desena o histogramă peste care va trasa curba distribuției normale.



	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
1	id	Numeric	4	0	Respondent ID Number	None	None	8	Right	Scale
2	wrkstat	Numeric	1	0	Labor Force Status	{0, NAP}...	0, 9	8	Right	Ordinal
3	marital	Numeric	1	0	Marital Status	{1, married}...	9	8	Right	Ordinal
4	agewed	Numeric	2	0	Age When First Married	{0, nap}...	0, 98, 99	8	Right	Ordinal
5	sibs	Numeric	2	0	Number of Brothers and Sisters	{98, dk}...	98, 99	8	Right	Ordinal
6	childs	Numeric	1	0	Number of Children	{8, Eight or ...	9	8	Right	Ordinal
7	age	Numeric	2	0	Age of Respondent	{98, DK}...	0, 98, 99	8	Right	Ordinal
8	birthmo	Numeric	2	0	Month in Which R Was Born	{0, NAP}...	0, 98, 99	8	Right	Ordinal
9	zodiac	Numeric	2	0	Respondents Astrological Sign	{0, NAP}...	0, 98, 99	8	Right	Ordinal
10	educ	Numeric	2	0	Highest Year of School Completed	{97, NAP}...	97, 98, 99	8	Right	Ordinal
11	degree	Numeric	1	0	R's Highest Degree	{0, Less tha...	7, 8, 9	8	Right	Ordinal
12	padeg	Numeric	1	0	Father's Highest Degree	{0, LT High ...	7, 8, 9	8	Right	Ordinal
13	madeg	Numeric	1	0					Right	Ordinal
14	sex	Numeric	1	0					Right	Ordinal
15	race	Numeric	1	0					Right	Ordinal
16	babies	Numeric	1	0					Right	Ordinal
17	preteen	Numeric	1	0					Right	Ordinal
18	teens	Numeric	1	0					Right	Ordinal
19	adults	Numeric	1	0					Right	Ordinal
20	income91	Numeric	2	0					Right	Ordinal
21	rincom91	Numeric	2	0					Right	Ordinal
22	region	Numeric	1	0					Right	Ordinal
23	xnorcsiz	Numeric	2	0					Right	Ordinal
24	srcbelt	Numeric	1	0					Right	Ordinal
25	size	Numeric	4	0					Right	Ordinal
26	partyid	Numeric	1	0	Political Party Affiliation	{0, Strong D...	8, 9	8	Right	Ordinal
27	vote92	Numeric	1	0	Voting in 1992 Election	{0, NAP}...	0, 8, 9	8	Right	Ordinal
28	polviews	Numeric	1	0	Think of Self as Liberal or Conservative	{0, NAP}...	0, 8, 9	8	Right	Ordinal
29	natenvir	Numeric	1	0	Improving & Protecting Environment	{0, NAP}...	0, 8, 9	8	Right	Ordinal
30	natheal	Numeric	1	0	Improving & Protecting Nation's Health	{0, NAP}...	0, 8, 9	8	Right	Ordinal
31	natcity	Numeric	1	0	Solving Problems of Big Cities	{0, NAP}...	0, 8, 9	8	Right	Ordinal
32	natcrime	Numeric	1	0	Halting Rising Crime Rate	{0, NAP}...	0, 8, 9	8	Right	Ordinal
33	natdrug	Numeric	1	0	Dealing With Drug Addiction	{0, NAP}...	0, 8, 9	8	Right	Ordinal
34	nateduc	Numeric	1	0	Improving Nation's Education System	{0, NAP}...	0, 8, 9	8	Right	Ordinal
35	cappun	Numeric	1	0	Favor or Oppose Death Penalty for Murder	{0, NAP}...	0, 8, 9	8	Right	Ordinal
36	gunlaw	Numeric	1	0	Favor or Oppose Gun Permits	{0, NAP}...	0, 8, 9	8	Right	Ordinal
37	grass	Numeric	1	0	Should Marijuana be Made Legal	{0, NAP}...	0, 8, 9	8	Right	Ordinal
38	relig	Numeric	1	0	Religious Preference	{1, Protesta...	8, 9	8	Right	Ordinal

La finalizarea comenzii, apăsați **Continue** apoi **OK**. Rezultatul este scris în **Output**.

Statistics

Number of Brothers and Sisters

N	Valid	1495,000
	Missing	5,000
Mean		3,714
Std. Error of Mean		,077
Median		3,000
Mode		2,000
Std. Deviation		2,981
Variance		8,887
Skewness		1,605
Std. Error of Skewness		,063
Kurtosis		3,710
Std. Error of Kurtosis		,126
Range		22,000
Minimum		,000
Maximum		22,000
Sum		5552,000
Percentiles	25	2,000
	50	3,000
	75	5,000

Explicații (pentru indicatorii care necesită așa ceva, restul îi găsiți în cursul de statistică)

1. Mean = media
2. Std. error of mean: eroarea standard a mediei. Se mai numește și abaterea standard a distribuției de eșantionare a mediei. Se calculează în cazul eșantioanelor aleatoare prin împărțirea abaterii standard (Std. Deviation) la rădăcina pătrată a numărului de cazuri:

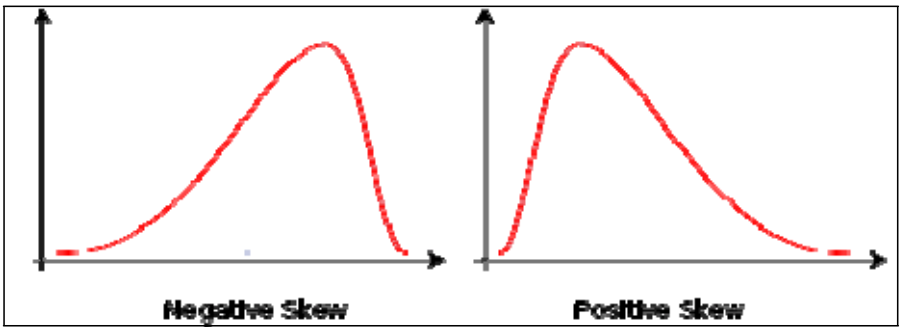
$$SE_{\bar{x}} = \frac{s}{\sqrt{n}}$$

Eroarea standard permite, de pildă calcularea marjei de eroare a estimării mediei din populație după formula:

Δ=1,96SE (la pragul de încredere de 95%)

3. Std. Deviation – abaterea standard
4. Skewness: este o măsură a asimetriei distribuției. Când indicele are valori pozitive, distribuția este alungită la dreapta. Kurtosis este o măsură a înălțimii distribuției. Pentru a stabili în funcție de aceste două măsuri gradul în care distribuția se abate de la

normalitate, valorile acestora se împart la eroarea standard corespunzătoare: dacă rezultatul raportării este mai mare de 2 atunci putem concluziona că distribuția se abate semnificativ de la ipoteza de normalitate. În cazul nostru: pentru skewness rezulta 25,4 iar pentru kurtosis 29,4 ambele valori fiind mult mai mari decât 2, iar ipoteza de normalitate se respinge. Histograma cu curbă normală de mai jos arată, de altfel, clar, că distribuția noastră nu este normală.



5. Amplitudinea, diferența dintre valoarea maximă și cea minimă, este de 22 de frați. Există în eșantion o persoană care raportează 22 de frați! Asta da familie!

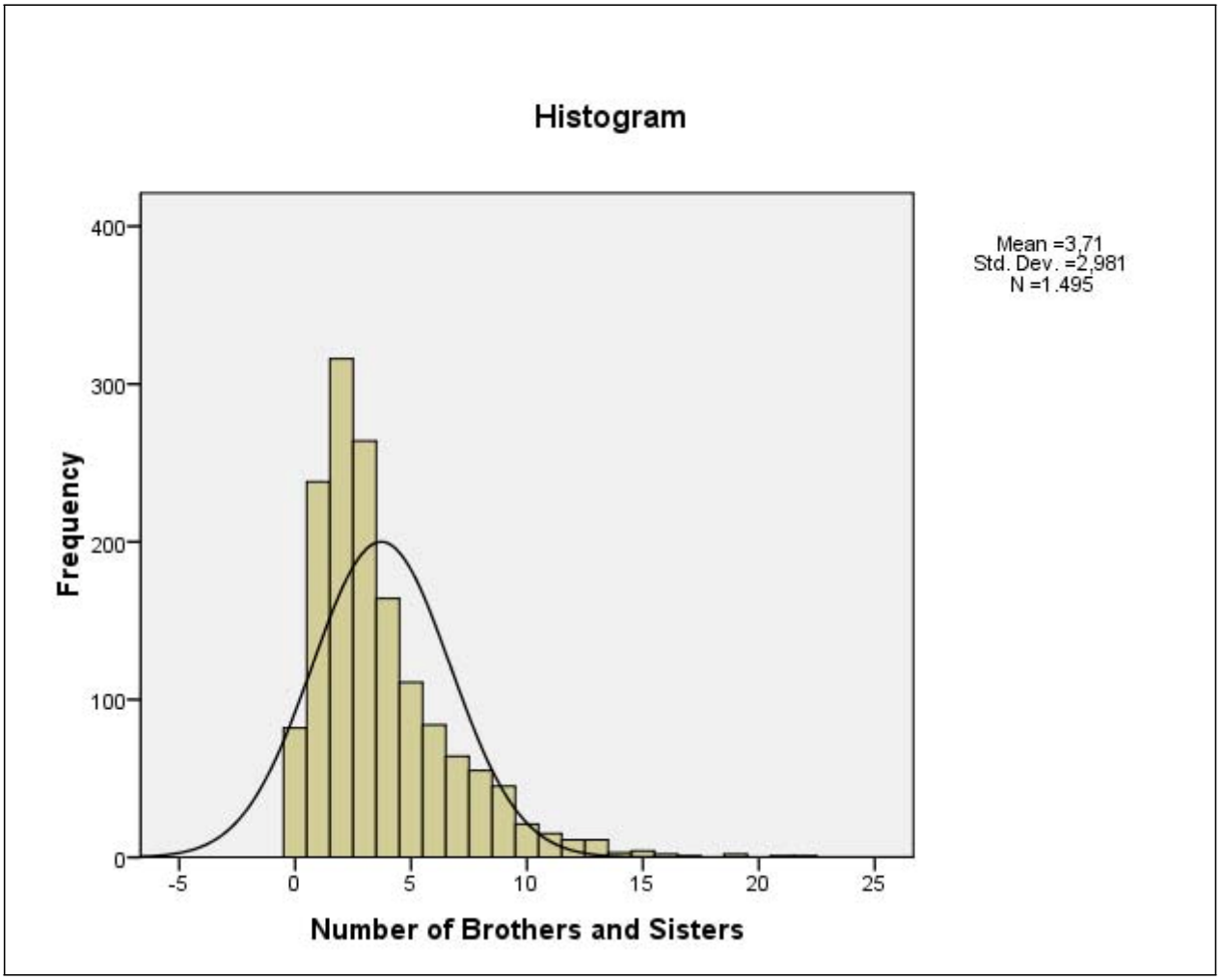
6. Percentilele 25, 50 și 75 sunt defapt cele 3 cuartile.

Number of Brothers and Sisters

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	82	5,5	5,5	5,5
	1	238	15,9	15,9	21,4
	2	316	21,1	21,1	42,5
	3	264	17,6	17,7	60,2
	4	164	10,9	11,0	71,2
	5	111	7,4	7,4	78,6
	6	84	5,6	5,6	84,2
	7	64	4,3	4,3	88,5
	8	55	3,7	3,7	92,2
	9	45	3,0	3,0	95,2
	10	21	1,4	1,4	96,6
	11	15	1,0	1,0	97,6
	12	11	,7	,7	98,3
	13	11	,7	,7	99,1
	14	3	,2	,2	99,3
	15	4	,3	,3	99,5
	16	2	,1	,1	99,7
	17	1	,1	,1	99,7
	19	2	,1	,1	99,9

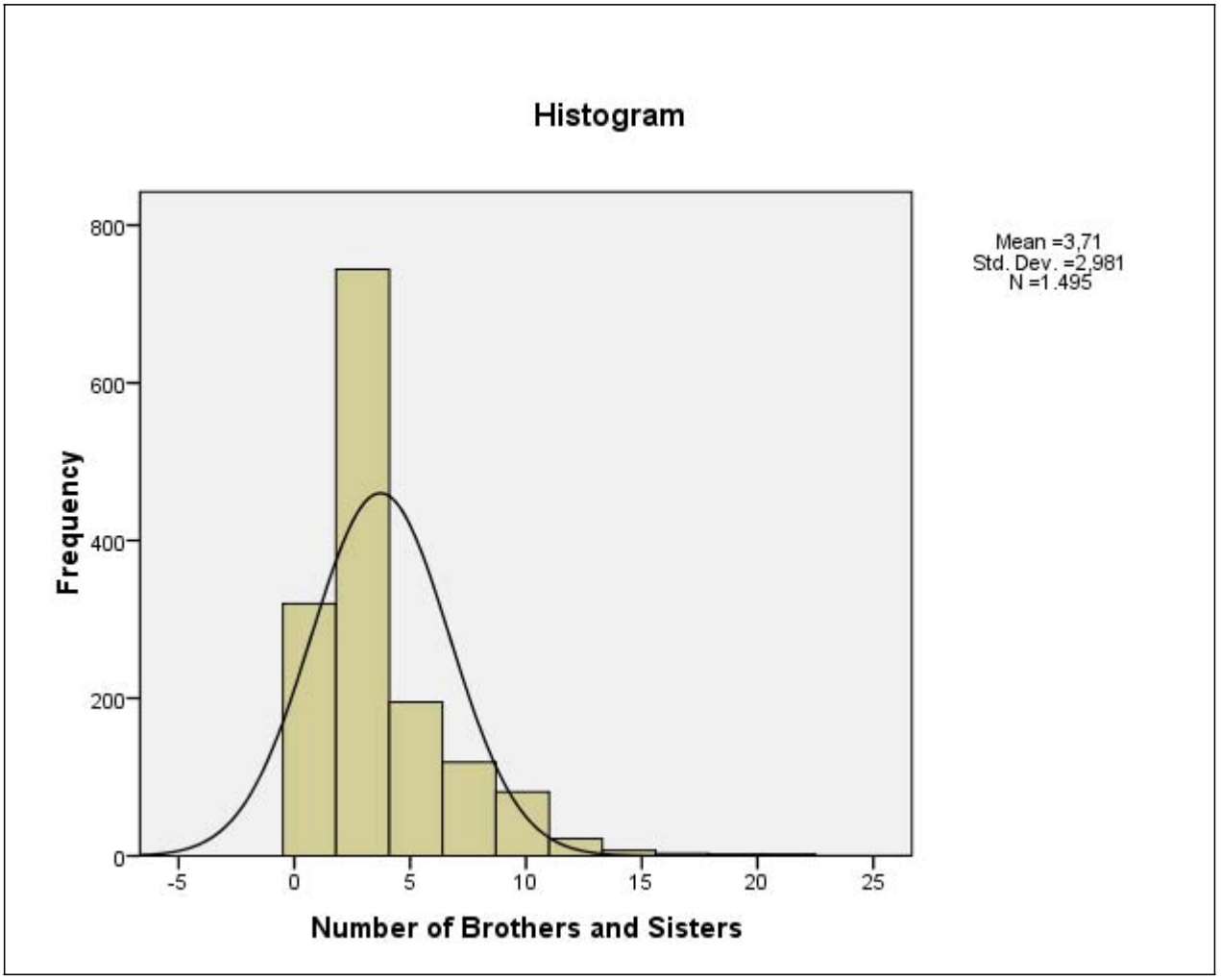
	21	1	,1	,1	99,9
	22	1	,1	,1	100,0
	Total	1495	99,7	100,0	
Missing	dk	2	,1		
	na	3	,2		
	Total	5	,3		
Total		1500	100,0		

Histograma, deși similară graficului cu bare, este diferită de aceasta. În timp ce la graficul cu bare, frecvențele relative sunt reflectate de diferențele dintre lungimile barelor, în cazul histogramelor, ponderea unui interval de valori de pe scală este reflectată în partea din suprafața de sub grafic acoperită de bara corespunzătoare intervalului respectiv.



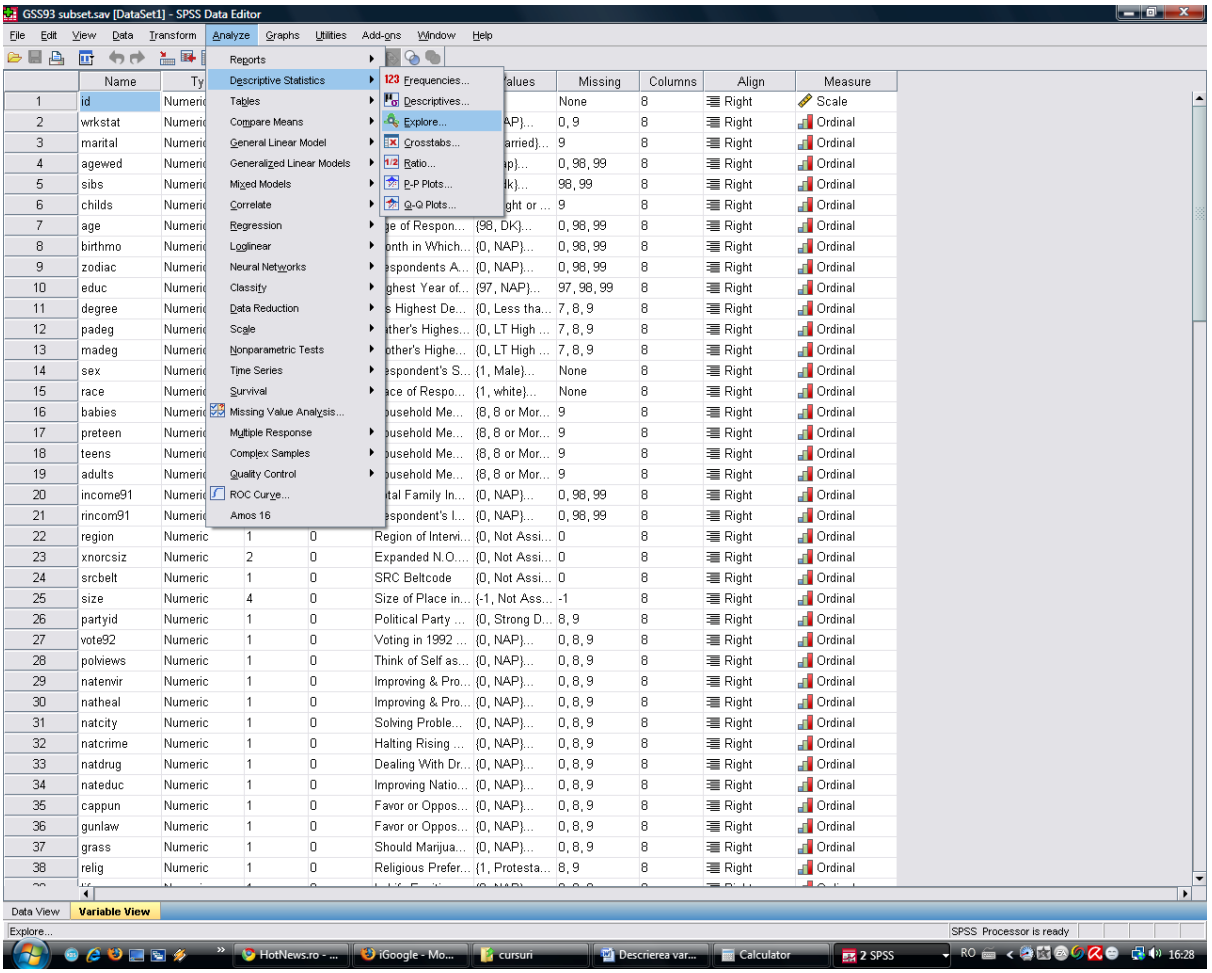
La fel ca și în cazul graficelor bară, histogramele pot fi editate. Diferența fundamentală dintre cele două tipuri de reprezentări grafice poate fi văzută comparând graficul original, mai sus, cu cel modificat, de mai jos, în care am schimbat intervalele de grupare alescalei variabilei.

Modificarea scalei duce la modificări mai mult sau mai puțin spectaculoase ale formei distribuției.

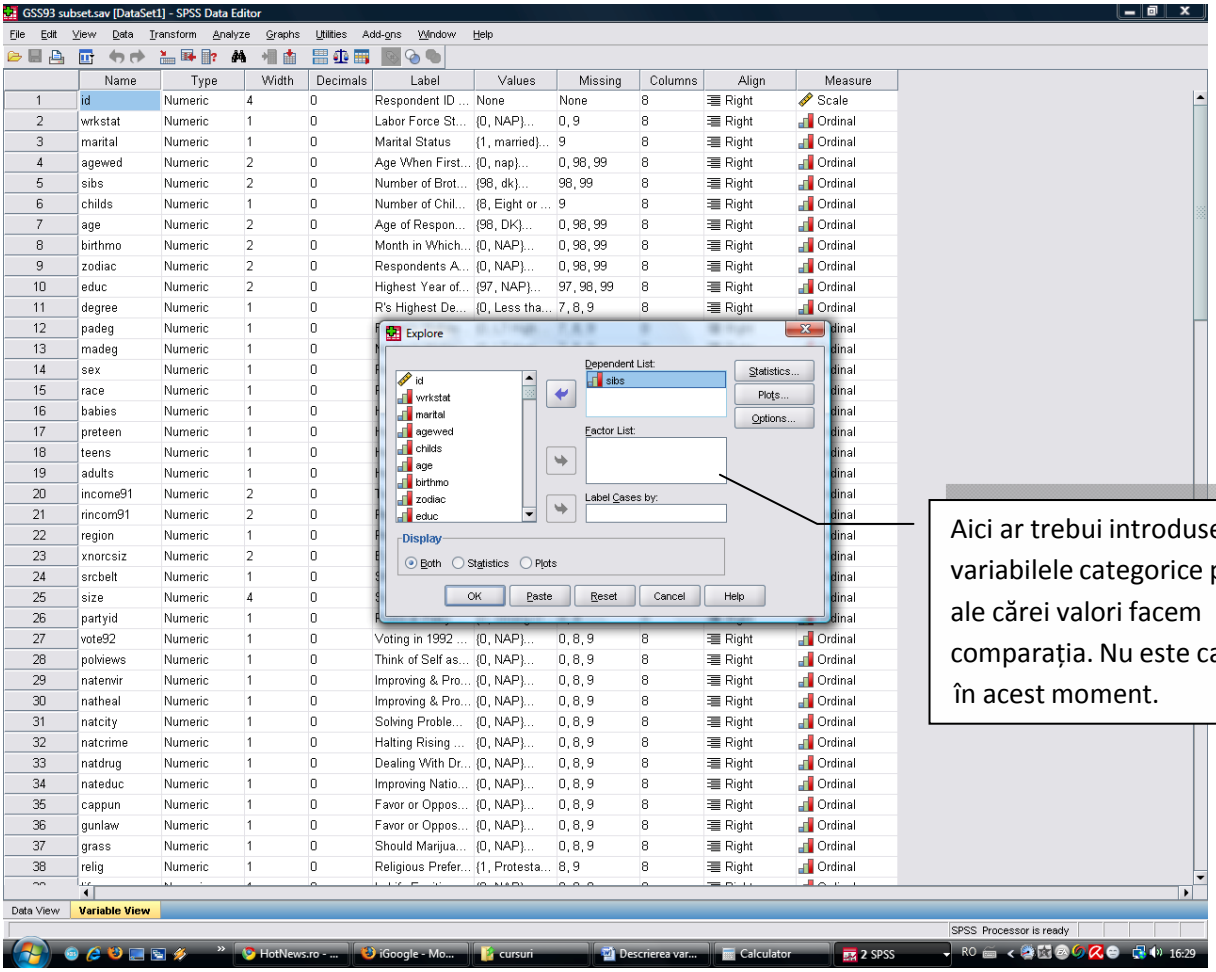


Soluții alternative pentru descrierea variabilelor numerice sunt la comenzile **Descriptives** și **Explore**. Întrucât opțiunea **Descriptives**, imediat următoare după **Frequencies**, oferă un număr redus de opțiuni, voi prezenta în detaliu comanda **Explore**

.



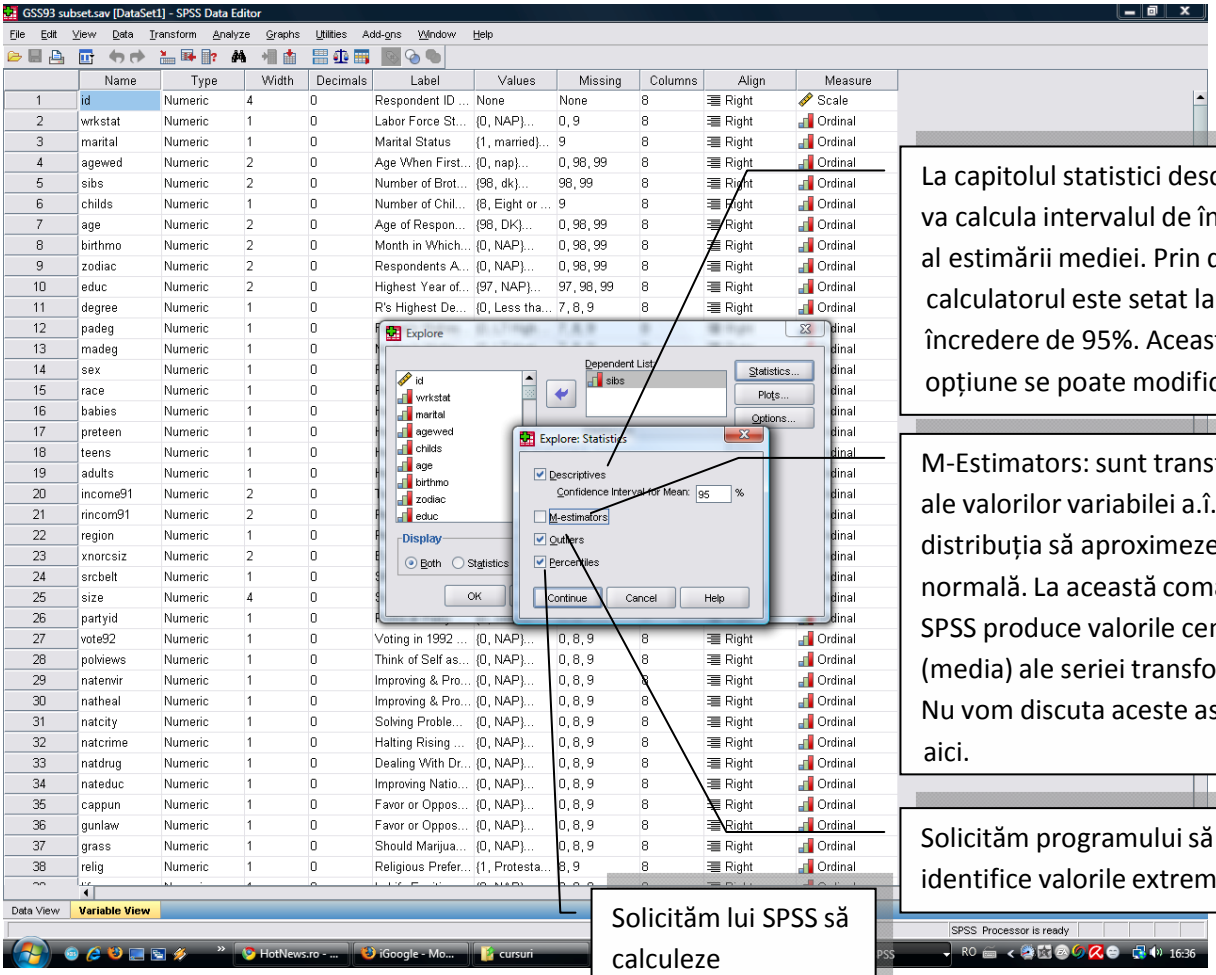
Comanda **Explore** este gândită pentru a analiza statistic şi vizual distribuția uneia sau mai multor variabile numerice pe categorii ale unor variabile categorice. Comanda se pretează foarte bine și pentru descrieri univariate.



Aici ar trebui introduse variabilele categorice pe ale cărei valori facem comparația. Nu este cazul în acest moment.

Pentru o descrierea satisfăcătoare vom activa toate opțiunile din dreapta ferestrei de dialog: **Statistics, Plot** și **Options**.

Apăsând butonul **Statistics** se deschide o fereastră de dialog cu patru opțiuni.



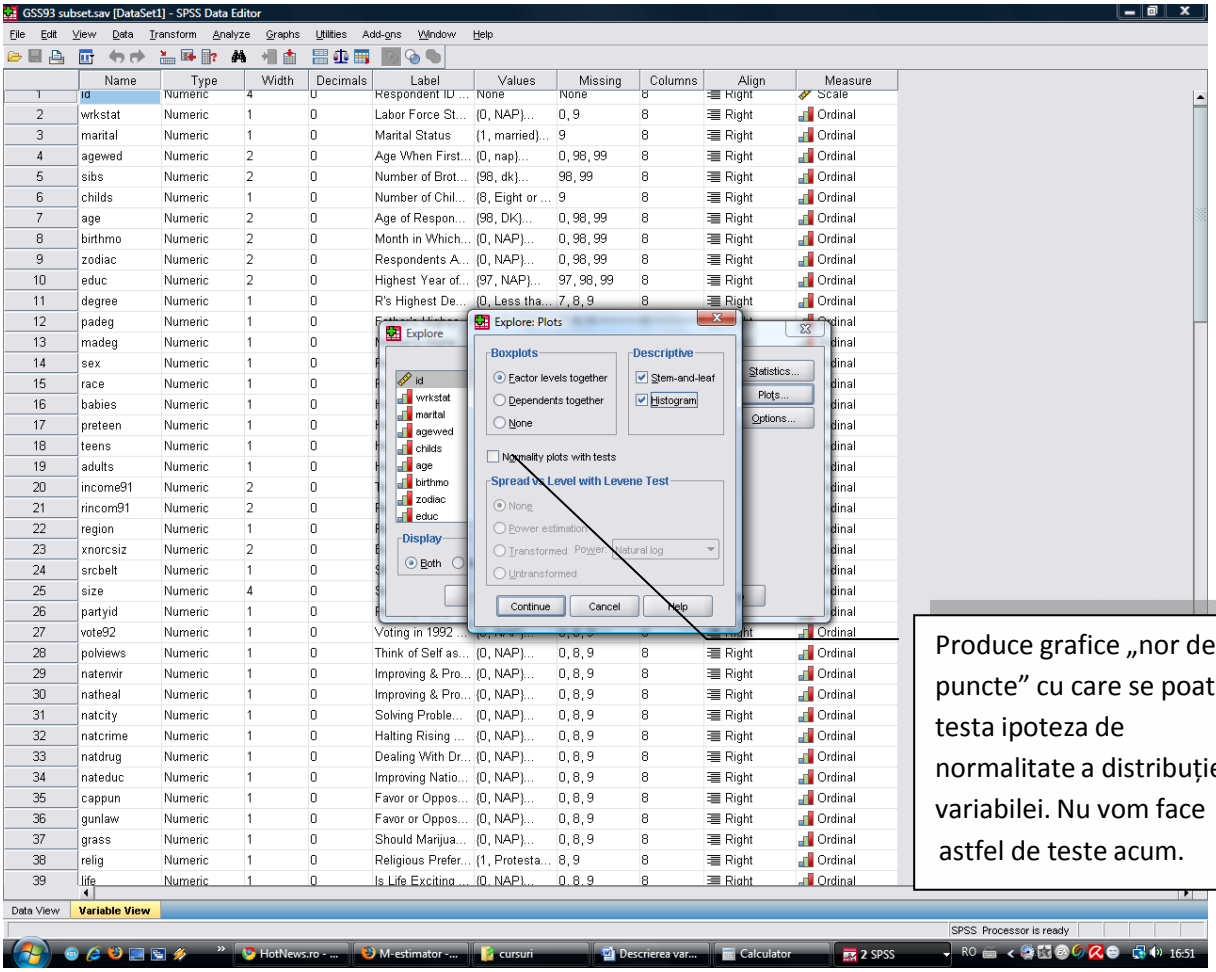
La capitolul statistici descriptive el va calcula intervalul de încredere al estimării mediei. Prin definiție, calculatorul este setat la pragul de încredere de 95%. Această opțiune se poate modifica.

M-Estimators: sunt transformări ale valorilor variabilei a.î. distribuția să aproximeze o curbă normală. La această comandă SPSS produce valorile centrale (media) ale seriei transformate. Nu vom discuta aceste aspecte aici.

Solicităm programului să identifice valorile extreme.

Solicităm lui SPSS să calculeze principalele valori percentilă.

După ce apăsăm **Continue**, deschidem opțiunea **Plots**.



Produce grafice „nor de puncte” cu care se poate testa ipoteza de normalitate a distribuției variabilei. Nu vom face astfel de teste acum.

Vom selecta trei tipuri de reprezentări grafice ale distribuției numărului de frați: box-plot (cutie), stem and leaf.

Apăsăm Continue apoi OK. Rezultatele sunt scrise în Output.

Rezultatele analizei cu Explore

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Number of Brothers and Sisters	1495	99,7%	5	,3%	1500	100,0%

O bună parte din rezultatele descriptive le cunoaștem deja din analiza anterioară. La rezultatele aceste se adaugă:

1. **Intervalele de încredere ale estimării mediei**, la pragul de încredere de 95%. Intervalul este cuprins între 3,56 și 3,86. Trebuie reamintit că acest interval este calculat plecând de la premisa că eșantionul este strict aleatoriu. Valorile limită ale intervalului de încredere pot diferi în cazul în care modalitatea de eșantionare aleatoare este alta (ex. stratificată, clusterizată etc., când se aplică alte formule de calcul a erorii standard a mediei)

2. **Media ajustată cu 5%:** se ordonează cazurile și se elimină cele mai mici 5% cazuri și cele mai mari 5% din cazuri. Cu cele 90% din cazuri rămase se calculează media obișnuită. Această medie se mai cheamă **medie robustă**, fiind eliminate cazurile extreme la care media este sensibilă. Observăm că media ajustată este mai mică decât cea obișnuită.

3. **Intervalul interquartilă (Interquartile range)** – diferența dintre valoarea quartilei 3 și valoarea quartilei 1. Este lungimea intervalului în care se întâlnesc cele 50% din cazuri cuprinse între cele două cuartile.

Descriptives

			Statistic	Std. Error
Number of Brothers and Sisters	Mean		3,71	,077
	95% Confidence Interval for Mean	Lower Bound	3,56	
		Upper Bound	3,86	
	5% Trimmed Mean		3,45	
	Median		3,00	

	Variance	8,887	
	Std. Deviation	2,981	
	Minimum	0	
	Maximum	22	
	Range	22	
	Interquartile Range	3	
	Skewness	1,605	,063
	Kurtosis	3,710	,126

Valorile percentile.

Percentiles

		Percentiles						
		5	10	25	50	75	90	95
Weighted Average(Definition 1)	Number of Brothers and Sisters	,00	1,00	2,00	3,00	5,00	8,00	9,00
Tukey's Hinges	Number of Brothers and Sisters			2,00	3,00	5,00		

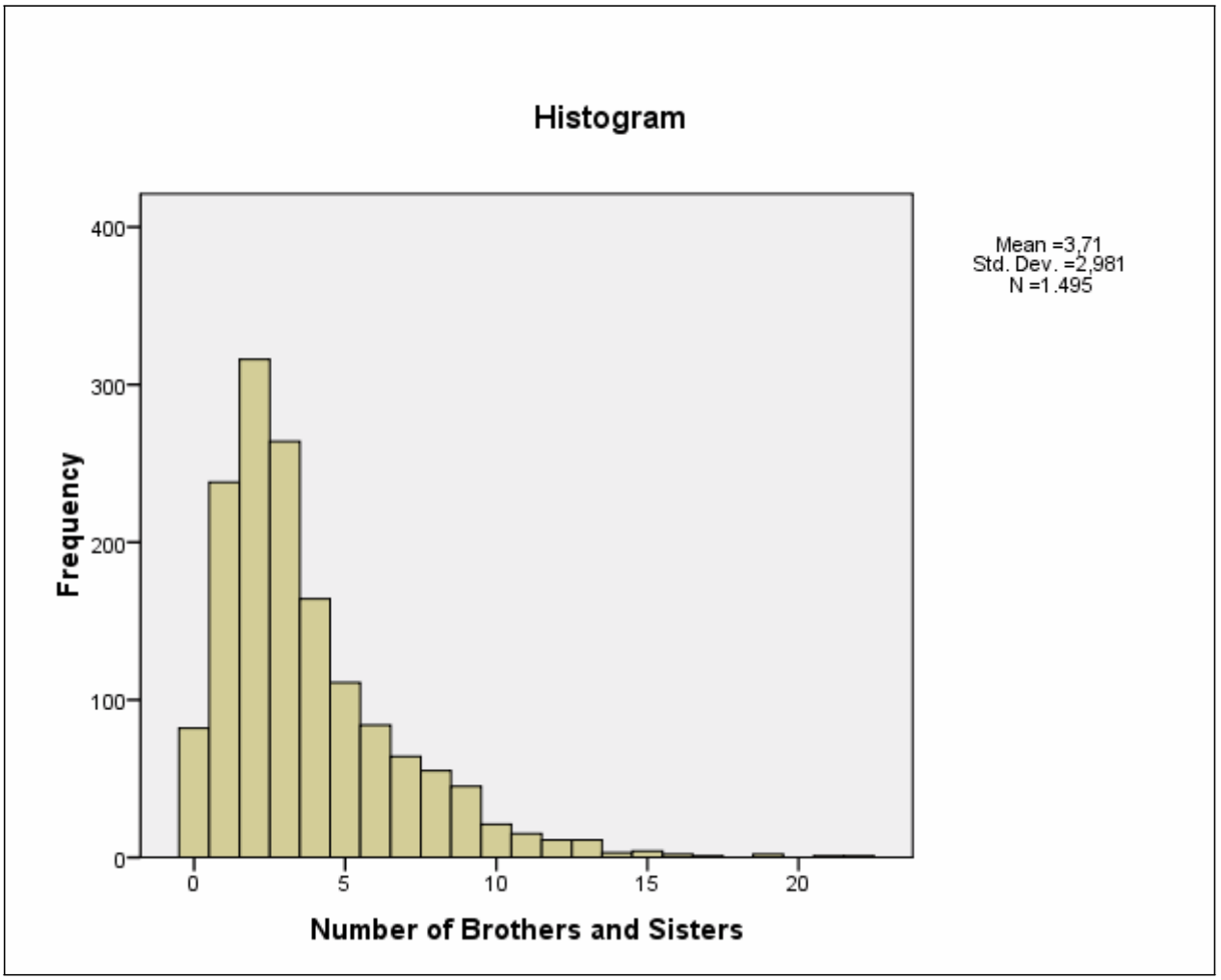
Pentru valorile extreme SPSS prezintă cele mai mici și cele mai mari seturi de 5 cazuri.

Extreme Values

			Case Number	Value
Number of Brothers and Sisters	Highest	1	743	22
		2	50	21
		3	121	19
		4	406	19
		5	693	17
	Lowest	1	1479	0
		2	1473	0
		3	1468	0
		4	1441	0
		5	1394	0[a]

a. Only a partial list of cases with the value 0 are shown in the table of lower extremes.

Number of Brothers and Sisters



Graficul stem and leaf (tulpină și frunză) este un mod de reprezentare potrivit pentru explorarea datelor numerice care păstrează atât forma distribuției (similară cu histograma) cât și informații referitoare la frecvențele valorilor în setul de date.

Stem= tulpină, înseamnă prima parte a numărului care exprimă valorile de pe scala variabilei. Frunza (leaf) este restul din număr, cu ajutorul cărora se exprimă frecvențele. Fiecare *stem and leaf* formează astfel un interval. Frecvența corespunzătoare acestui interval este scrisă la coloana Frequency.

În cazul nostru, numărul de valori fiind relativ mic, se consideră tulpină valoarea în întregime iar frunzele, cu care se reprezintă și numărul de cazuri, vor fi alcătuite din zerouri. Fiecare frunză (zero) reprezintă câte 4 cazuri.

Number of Brothers and Sisters Stem-and-Leaf Plot

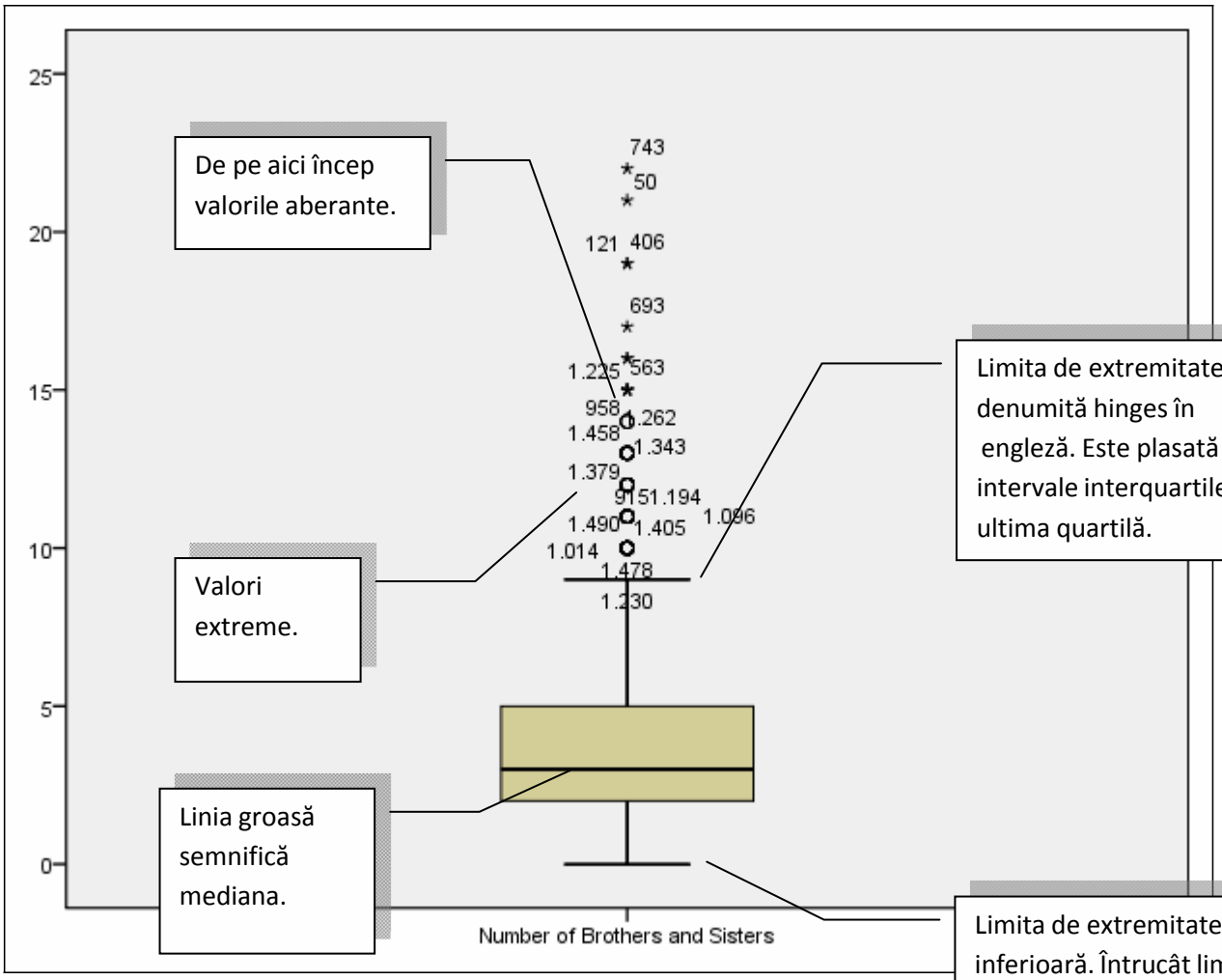
Frequency Stem & Leaf

```
 82,00      0 .  00000000000000000000
  ,00       0 .
238,00      1 .  00000000000000000000000000000000000000000000000000000000000000
  ,00       1 .
316,00      2 .  00000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000
  ,00       2 .
264,00      3 .  000000000000000000000000000000000000000000000000000000000000000000
  ,00       3 .
164,00      4 .  0000000000000000000000000000000000000000
  ,00       4 .
111,00      5 .  00000000000000000000000000000
  ,00       5 .
 84,00      6 .  000000000000000000000
  ,00       6 .
 64,00      7 .  0000000000000000
  ,00       7 .
 55,00      8 .  00000000000000
  ,00       8 .
 45,00      9 .  00000000000
 72,00 Extremes    (>=10)

Stem width:   1
Each leaf:      4 case(s)
```

Graficul cutie (box-plot)

Limitele cutiei sunt trasate la valorile cuartilei 1 și 3. Linia groasă din interiorul cutiei reprezintă mediana. Liniile de deasupra și dedesubt de cutiei marchează limitele de la care valorile sunt considerate extreme. Aceste limite sunt trasate la 1,5 intervale interquatile de limitele cutiei (cuartile). Prin convenție, valorile care se află la distanță mai mare de 3 intervale interquartile de cuartilele 1 și 3 sunt considerate aberante. De pildă , cazul 1405 are valoare extremă, în timp ce cazul 693 este aberant.



De pe aici încep valorile aberante.

Limita de extremitate. Este denumită hinges în engleză. Este plasată la 1,5 intervale interquartile de ultima quartilă.

Valori extreme.

Linia groasă semnifică mediana.

Limita de extremitate inferioară. Întrucât limita s-ar afla sub valoarea minimă a variabile, SPSS îl trunchiază.

Teste non-parametrice

Se folosesc pentru a testa ipoteze privind distribuția unor variabile categorice (sau non-parametrice)

Testul binomial

Problemă: să se testeze ipoteza că distribuția pe sexe din baza de date droguri2000.sav nu se abate semnificativ de la distribuția teoretică: 52% fete la 48% băieți.

Soluție

Din punct de vedere logic, problema poate fi reformulată astfel: care este probabilitatea de a obține distribuția pe sexe din baza de date prin selecție aleatoare dintr-o populație în care fetele reprezintă 52% iar băieții reprezintă 48%?

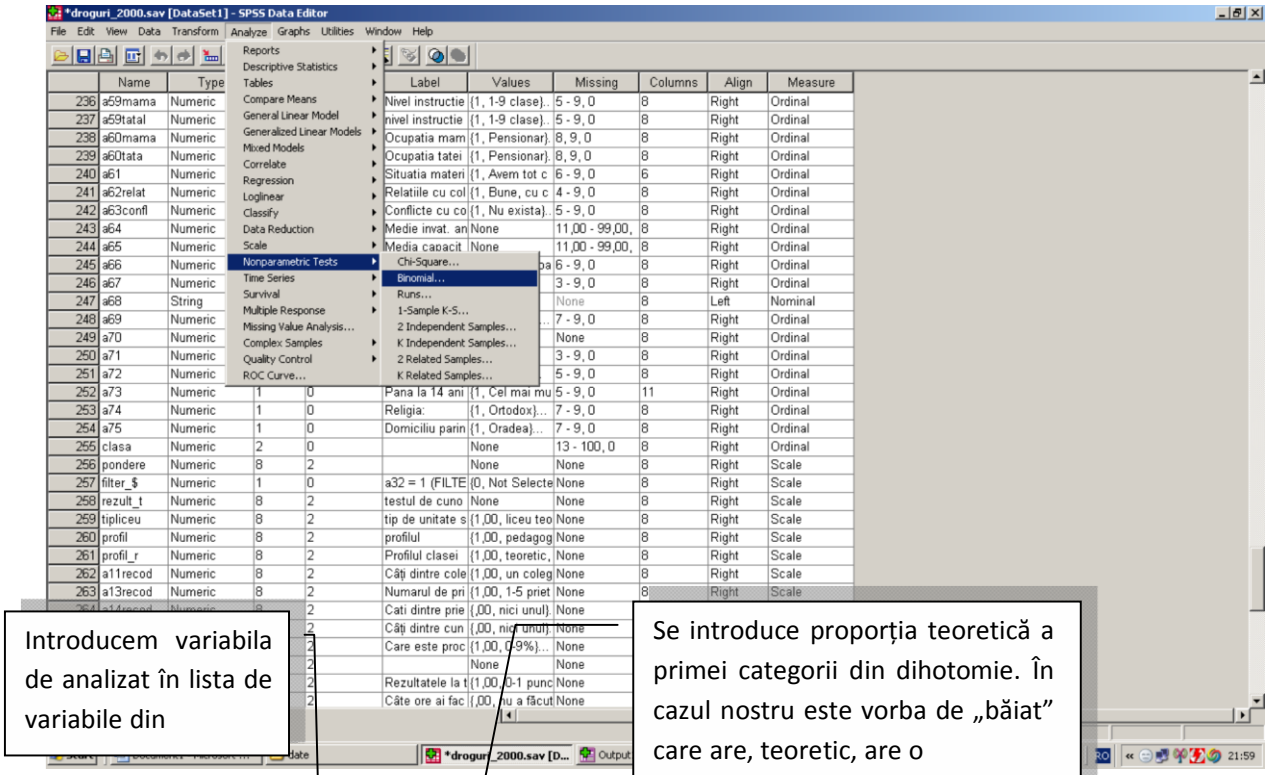
Această problemă se poate rezolva cu testul binomial care are următoarea formulă (copiată de pe net ☺):

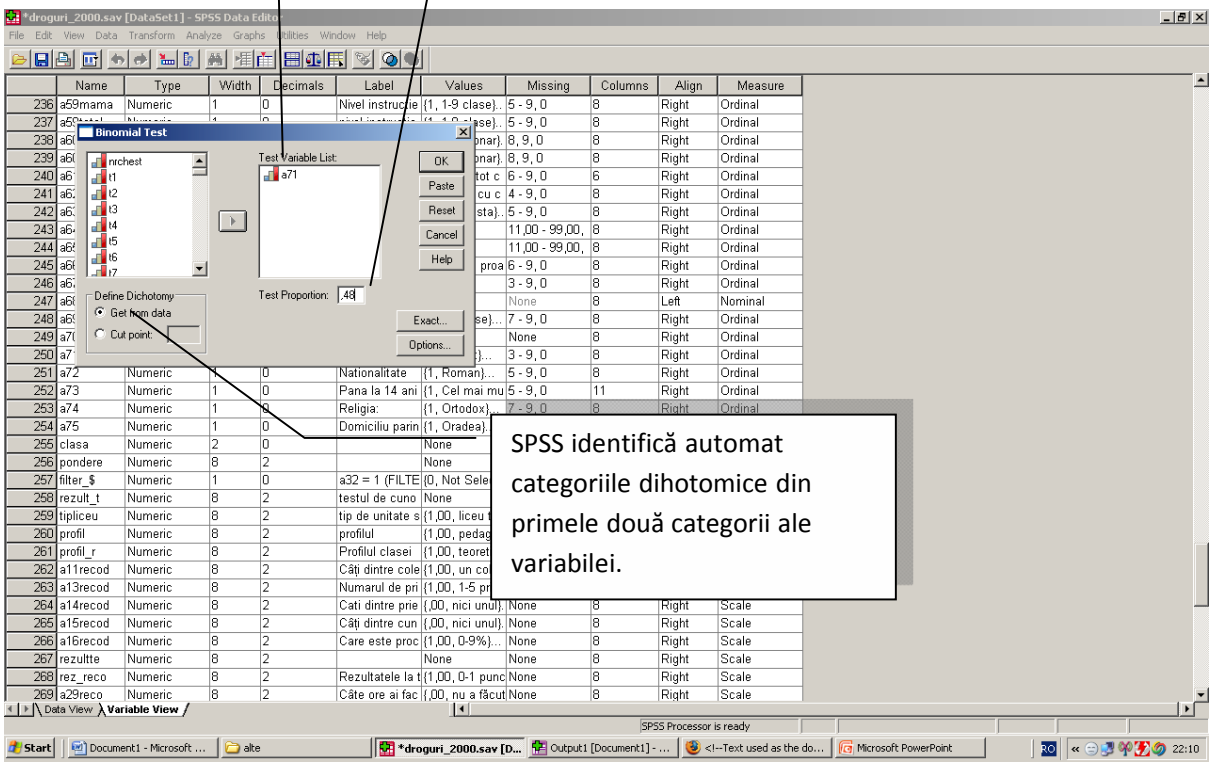
$$P\left(k \text{ successes in } n \text{ trials}\right) = \binom{n}{k} p^k q^{n-k}$$

Probabilitatea de a obține k evenimente din n încercări se calculează ca produs al: combinațiilor de n luate câte k multiplicat cu probabilitatea teoretică a succesului ridicată la numărul de succese înmulțit cu probabilitatea teoretică a eşecului ridicată la puterea numărului de eşecuri. În exemplul nostru, numărul de încercări este de 978 iar evenimentele sunt: a fi băiat (succes ☺) şi fată (eşec, ☺). Numărul de succese este de 435 iar cel de eşecuri 543. Aplicabilitatea procedurii este limitată la variabile dihotomice ori politomice dihotomizate.

Potrivit convențiilor utilizate în statistică, distribuția empirică va fi considerată semnificativ diferită de cea teoretică dacă probabilitatea calculată folosind formula de mai sus va fi mai mică de 0,05.

Pentru a calcula această probabilitate folosim procedura Binomial din Non-parametric.


*droguri_2000.sav [DataSet1] - SPSS Data Editor
Analyze
Nonparametric Tests
Binomial...
Introducem variabila de analizat în lista de variabile din
Se introduce proporția teoretică a primei categorii din dihotomie. În cazul nostru este vorba de „băiat” care are, teoretic, are o probabilitate de 0,48 (sau 48%)


Binomial Test
Test Variable List:
a71
Test Proportion: .48
Define Dichotomy
Get from data
Cut point:
OK
Paste
Reset
Cancel
Help
Exact...
Options...
SPSS identifică automat categoriile dihotomice din primele două categorii ale variabilei.

Rezultatul produs de procedură este simplu de interpretat.

O procedură de aproximare a unei distribuții normale pentru o funcție binomială.

Binomial Test

		Category	N	Observed Prop.	Test Prop.	Asymp. Sig. (1-tailed)
Sex	Group 1	Barbat	435	,44	,48	,015(a,b)
	Group 2	Femeie	543	,56		
	Total		978	1,00		

a Alternative hypothesis states that the proportion of cases in the first group < ,48.
b Based on Z Approximation.

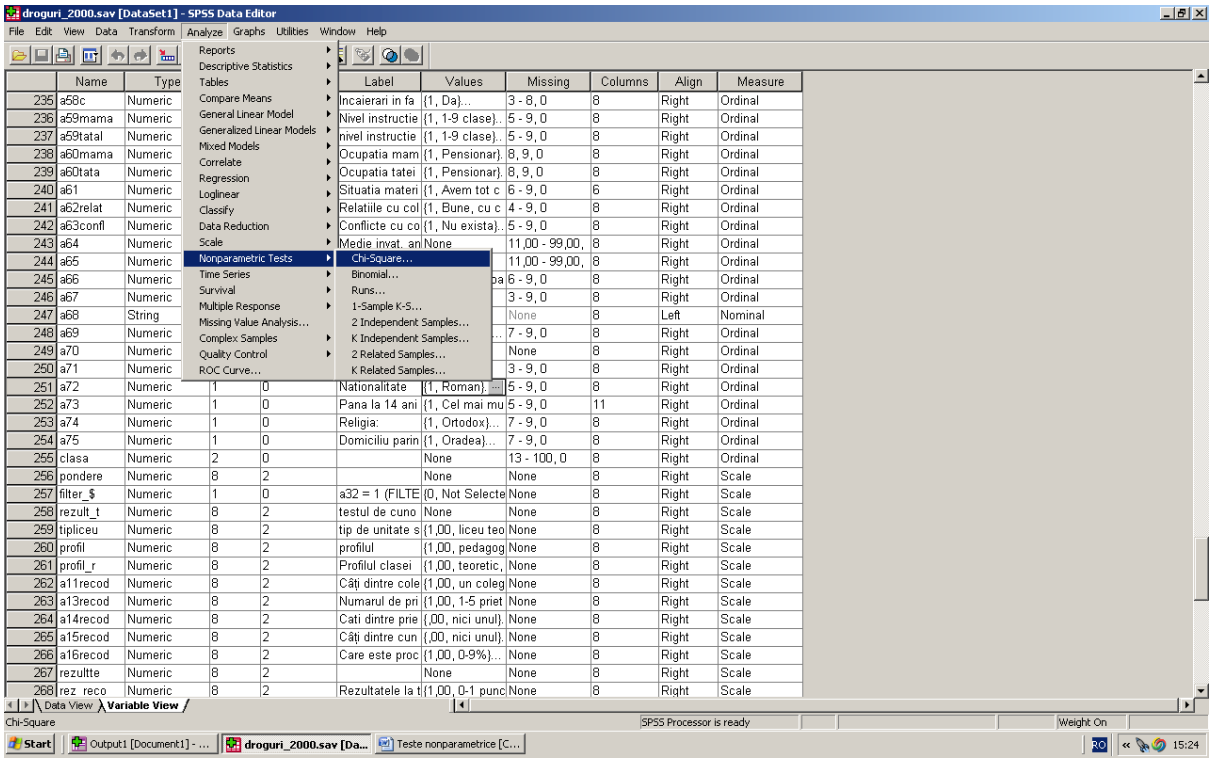
Rezultatul arată că probabilitatea de a obține distribuția empirică dintr-o populație cu distribuția teoretică dată în problemă este de 0,015. Această probabilitate fiind mai mică de 0,05, se respinge ipoteza nulă, a egalității distribuției teoretice şi a celei empirice şi se admite ipoteza alternativă că proporția bărbaților în populație este mai mică de 0,48.

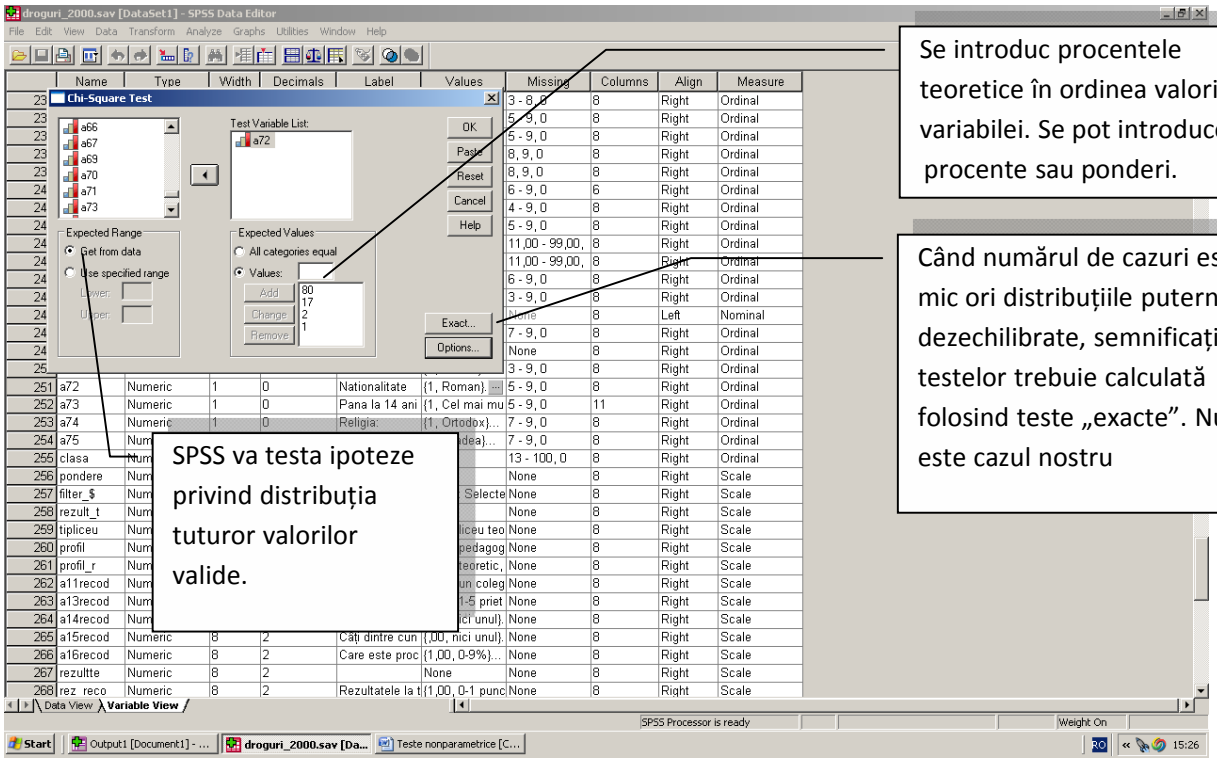
Testul chi-pătrat

Problemă: distribuția pe etnii din baza de date droguri2000.sav corespunde distribuției teoretice: români 80%, maghiari 17%, romi 2%, alta 1%?

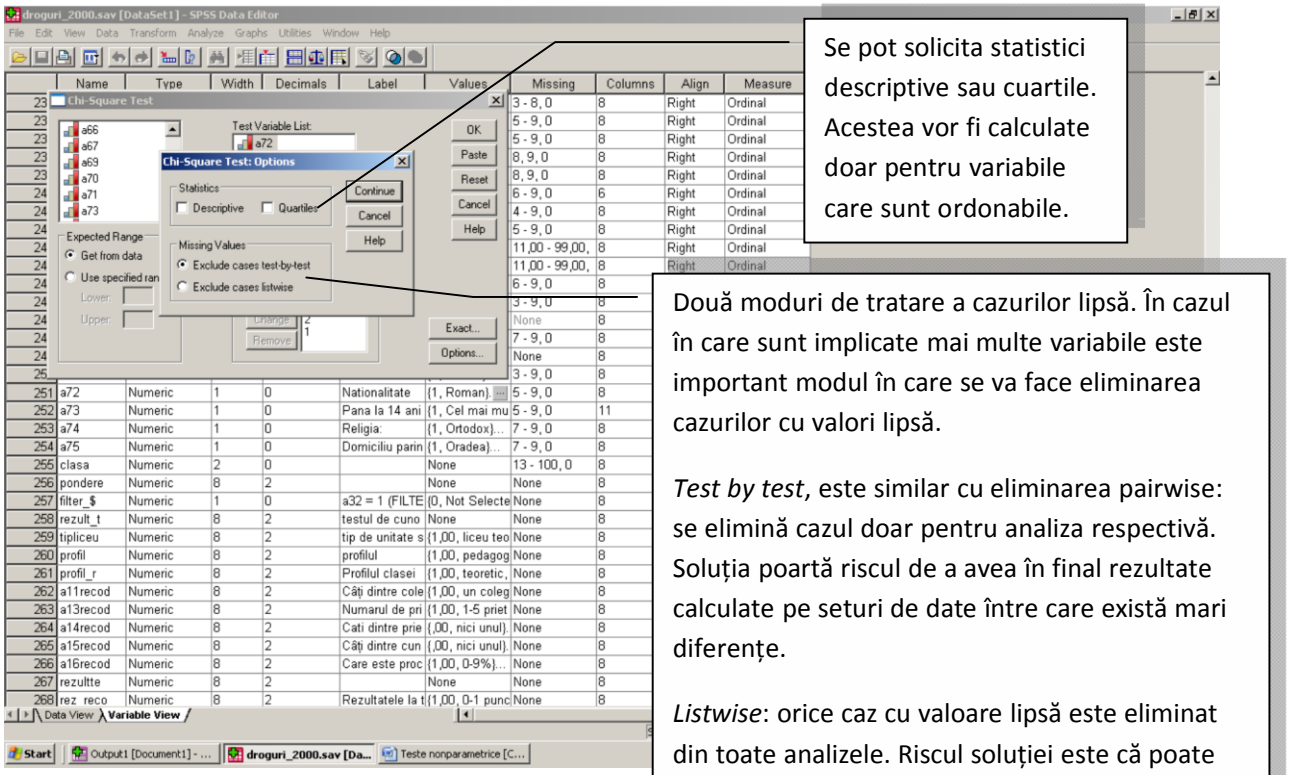
Este evident că procedura pe care urmează să o introducem este utilă pentru testarea reprezentativității unor eşantioane în funcție de distribuțiile variabilelor categorice sau, pur şi simplu, dacă dorim să testăm ipoteze privind distribuția unei variabile categorice.

Soluția acestei probleme constă în testarea ipotezei cu ajutorul testului chi-pătrat univariat.


droguri_2000.sav [DataSet1] - SPSS Data Editor
File Edit View Data Transform Analyze Graphs Utilities Window Help
Reports
Descriptive Statistics
Tables
Compare Means
General Linear Model
Generalized Linear Models
Mixed Models
Correlate
Regression
Loglinear
Classify
Data Reduction
Scale
Nonparametric Tests
Time Series
Survival
Multiple Response
Missing Value Analysis...
Complex Samples
Quality Control
ROC Curve...
Chi-Square...
Binomial...
Runs...
1-Sample K-S...
2 Independent Samples...
K Independent Samples...
2 Related Samples...
K Related Samples...
Name Type Width Decimals Label Values Missing Columns Align Measure
Data View Variable View
Chi-Square
SPSS Processor is ready
Weight On
Start
Output1 [Document1] - ...
droguri_2000.sav [Da...
Teste nonparametrice [C...
15:24


droguri_2000.sav [DataSet1] - SPSS Data Editor
Chi-Square Test
Test Variable List:
a72
OK
Paste
Reset
Cancel
Help
Expected Range
Get from data
Use specified range
Lower:
Upper:
Expected Values
All categories equal
Values:
Add
Change
Remove
80
17
2
1
Exact...
Options...
SPSS va testa ipoteze privind distribuția tuturor valorilor valide.
Se introduc procentele teoretice în ordinea valorilor variabilei. Se pot introduce procente sau ponderi.
Când numărul de cazuri este mic ori distribuțiile puternic dezechilibrate, semnificația testelor trebuie calculată folosind teste „exacte”. Nu este cazul nostru
Data View Variable View
SPSS Processor is ready
Weight On
Start
Output1 [Document1] - ...
droguri_2000.sav [Da...
Teste nonparametrice [C...
15:26

Dacă apăsăm butonul Options vom avea de ales între câteva opțiuni statistice neobligatorii pentru finalizarea testului chi-pătrat.



Formula lui Chi-pătrat este:

Chi-pătrat=$\sum (t-o)^2/t$

Unde:

T – frecvențele teoretice

O – frecvențele observate

După ce apăsăm pe OK obținem următorul rezultat în fereastra Output (aici este copiat din html exportat din rezultatul original).

Chi-Square Test

Frecvențele observate.

Frecvențele teoretice corespunzătoare procentelor testate.

Frequencies

Nationalitate

	Observed N	Expected N	Residual
Roman	801	782,4	18,6
Maghiar	168	166,3	1,7
Rrom	2	19,6	-17,6
Alta	7	9,8	-2,8
Total	978		

Diferența dintre frecvențele observate şi cele teoretice

Mărimea testului chi-pătrat

Test Statistics

	Nationalitate
Chi-Square(a)	17,015
df	3
Asymp. Sig.	,001

a 0 cells (,0%) have expected frequencies less than 5. The minimum expected cell frequency is 9,8.

Numărul de grade de libertate

Probabilitatea testului chi-pătrat.

Întrucât probabilitatea testului chi-pătrat este mai mică de 0,05 respingem ipoteza nulă şi o acceptăm pe cea alternativă: distribuția observată pe etnii nu corespunde celei teoretice.

Asocierea. Testul Chi-pătrat

Asocierea se testează pentru a verifica dacă există relație (asociere) între variabile categorice.

Problemă: Există asociere între sexul elevilor şi fumat?

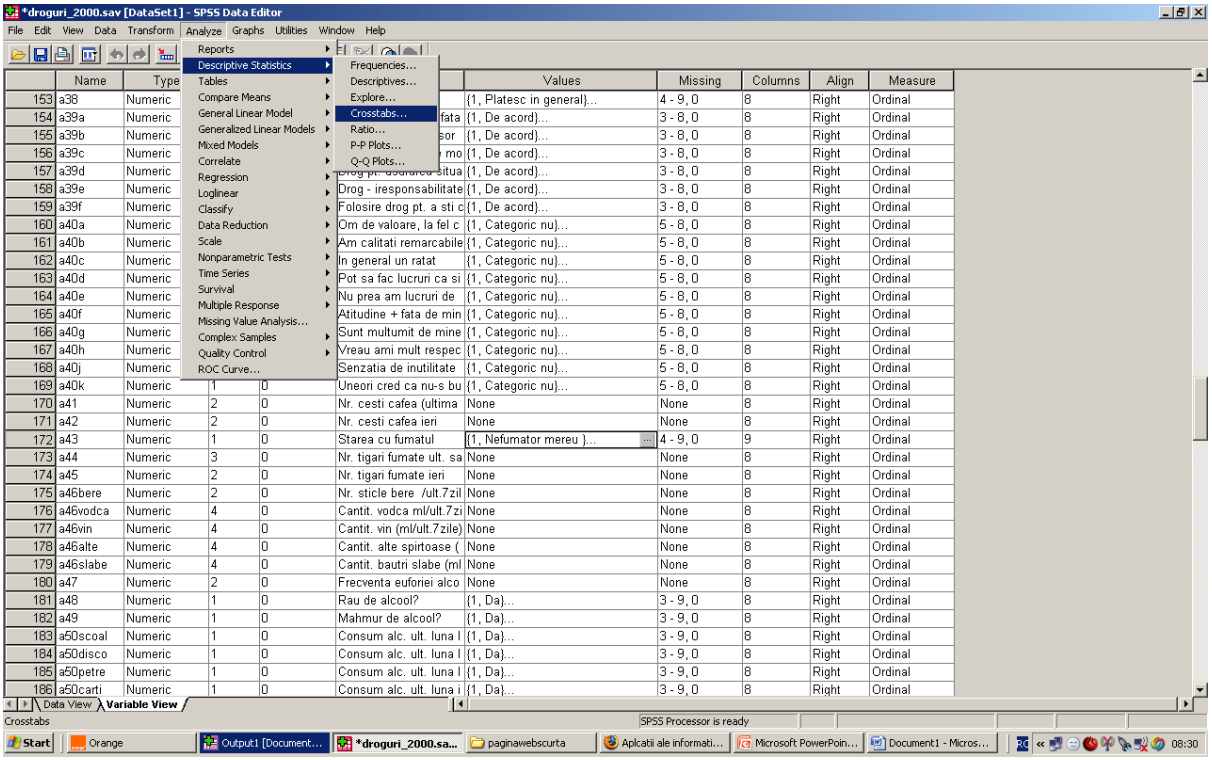
Cele două variabile fiind categorice vom fi nevoiți să rezolvăm problema prin analize de asociere. Vom dezvolta tema în trei etape:

- Realizarea tabelului de asociere
- Identificarea modelului de asociere prin inspectare vizuală a tabelului de asociere
- Testarea statistică a asocierii

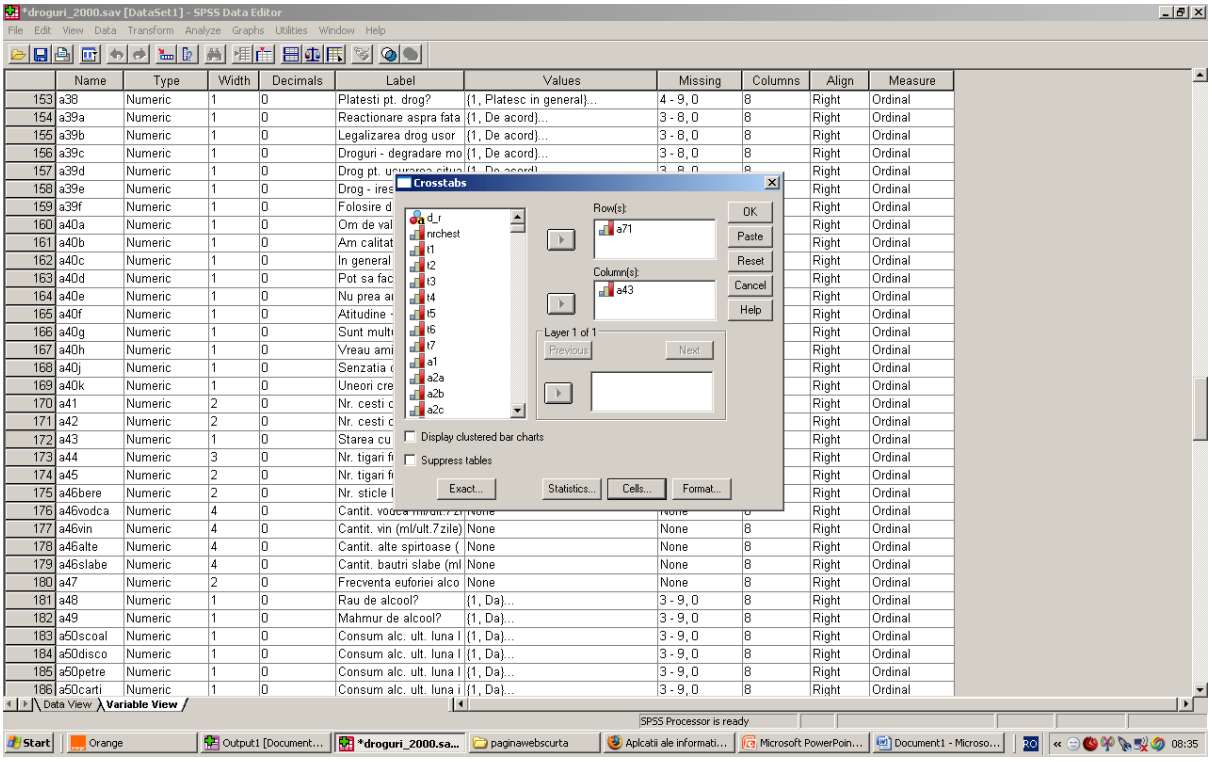
Realizarea tabelului de asociere (contingență)

Tabelul de asociere este un tabel cu dublă intrare în care pe linii se trec valorile variabilei anterioare (posibil cauză) iar pe coloane valorile variabilei subsecvente (posibil efect). Tabelul de asociere se realizează pentru scop de analiză din meniul Descriptives iar pentru scop de raportare recomand folosirea meniului Custom Tables.

Realizarea unui tabel de asociere simplu cu comanda crosstabs



Denumirea englezească a tabelelor de asociere este Crosstabs.



Pentru a obține tabele de asociere simple vom introduce la linii (rows) variabila gen (anterioară celeilalte) iar pe coloane variabile variabila a43 (Starea cu fumatul). Fără a mai selecta vreo opțiune, apăsăm pe OK. Rezultatul, exportat în html este prezentat mai jos:

```
[DataSet1] C:\spatiul de lucru\voyager\paginawebscurta\date\droguri_2000.sav
```

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Sex * Starea cu fumatul	955	95,5%	45	4,5%	1000	100,0%

Frecvențe interioare. În baza de date avem 134 de băieți care declară fumători.

Sex * Starea cu fumatul Crosstabulation

Count

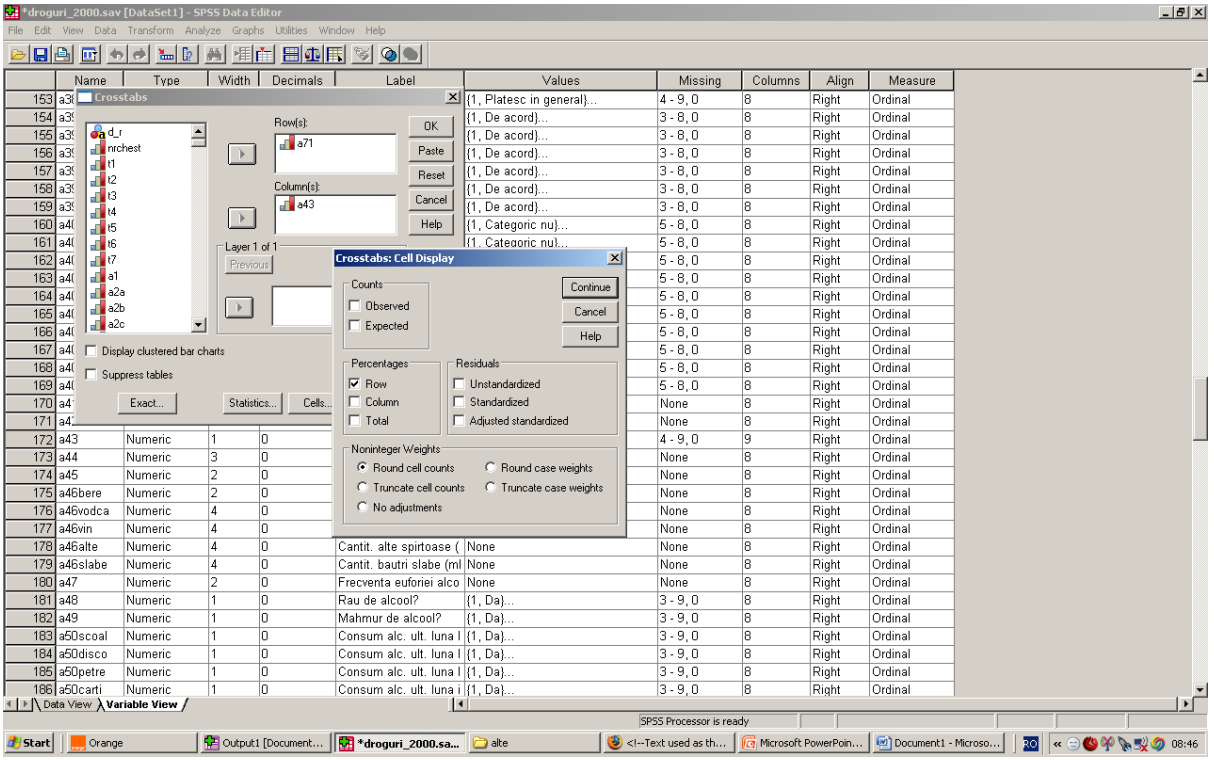
		Starea cu fumatul			Total
		Nefumator mereu	Am incercat doar	Sunt/am fost fumator	Nefumator mereu
Sex	Barbat	121	171	134	426
	Femeie	183	230	116	529
Total		304	401	250	955

Frecvențe marginale (total pe linii). Numărul total de băieți care au dat răspunsuri valide la întrebare este de 426.

Identificarea asocierilor folosind procentele pe linii

Frecvențele interioare sunt, cel mai adesea, inutile când se încearcă identificarea asocierii din variabile. O procedură care poate ajuta la stabilirea aproximativă a existenței unei legături între variabile categorice constă în compararea procentelor pe linii cu procentele de pe linia totalurilor. Aceasta se poate realiza tot în meniul crosstabs sau în meniul Custom Tables.

Având deschisă fereastra de dialog a meniului crosstabs apăsăm butonul Cells (dreapta jos). În fereastra de dialog selectăm opțiunea Row (de la capitolul percentages) şi deselectăm opțiunea Observed (de la Counts). Deselectarea fercvențelor observate va ajuta la lectura rezultatelor.



Pentru executarea comenzii, apăsăm Continue apoi OK. Rezultatul, în html, este prezentat mai jos:

```
[DataSet1] C:\spatiul de lucru\voyager\paginawebscurta\date\droguri_2000.sav
```

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Sex * Starea cu fumatul	955	95,5%	45	4,5%	1000	100,0%

Mult mai multe fete se declară nefu mătoare.

Linia totalurilor

Sex * Starea cu fumatul Crosstabulation
% within Sex

		Starea cu fumatul			Total
		Nefumator mereu	Am incercat doar	Sunt/am fost fumator	Nefumator mereu
Sex	Barbat	28,4%	40,1%	31,5%	100,0%
	Femeie	34,6%	43,5%	21,9%	100,0%
Total		31,8%	42,0%	26,2%	100,0%

Procentul băieților fumători este cu aproape 10 puncte procentuale mai mare decât al fetelor.

Intuitiv, ştim că în cazul în care nu ar exista asociere, procentele de pe linii din căsuțele interioare ar trebui să fie asemănătoare, între ele şi apropiate de cele de pe linia totalurilor. Observăm că acest lucru

este valabil pentru procentul celor care spun că au încercat doar să fumeze (între 40 şi 43,5%). În schimb, pentru celelalte categorii ale variabilei Starea cu fumatul avem diferențe mult mai mari: procentul fetelor care declară că sunt nefumătoare este cu aproape 16,2% mai mare decât al băieților aflați în situație similară în timp ce procentul băieților care se declară fumători este cu 9,6% mai mare decât în cazul fetelor. Din aceste comparații simple rezultă că între cele două variabile există o anumită legătură, băieții având o şansă mai mare decât fetele să de declare fumători.

Testarea statistică a asocierii

Adeseori însă nici procentele pe linii nu sunt suficiente pentru a identifica relații de asociere. În orice caz, stabilirea fermă a unei astfel de relații nu se poate baza doar pe inspecție vizuală ci trebuie să aibă la bază testul statistic al asocierii. Cel mai important astfel de test, aplicabil în aproape toate cazurile de asociere – indiferent de numărul de categorii al variabilelor introduse, de calitatea ordinală sau nu a acestora – este testul chi-pătrat. Testul acesta compară statistic frecvențele interioare observate cu frecvențele teoretice, corespunzătoare ipotezei de independență. De ex. pe noi ne interesează dacă există o relație între categoria de gen a adolescenților din eşantion şi fumat? Pentru a rezolva această problemă va trebui să calculăm frecvențele interioare pentru ipoteza de independență a celor două variabile.

- Plecăm de la probabilități: am învățat că două evenimente A şi B sunt independente dacă P(A şi B)=P(A)P(B)
- Aplicând formula la tabelul de mai sus, evenimentul de a fi băiat şi de a se declara fumător sunt independente dacă probabilitatea evenimentului conjugat este egală cu produsele probabilităților celor două evenimente separate.
    - P(Băiat)=0,446
    - P(fumător)=0,262
    - P(Băiat fumător)=0,446X0,262=0,117, ceea ce este mai puțin decât probabilitatea empirică, de 0,14
- Frecvența teoretică se obține înmulțind probabilitatea teoretică cu numărul de cazuri: 0,117X955=111,7
- Acest calcul se repetă pentru fiecare căsuță a tabelului de asociere.

Măsura abaterii frecvențelor empirice de cele teoretice se obține aplicând formula lui chi-pătrat:

Chi-pătrat=$\sum (t-o)^2/t$

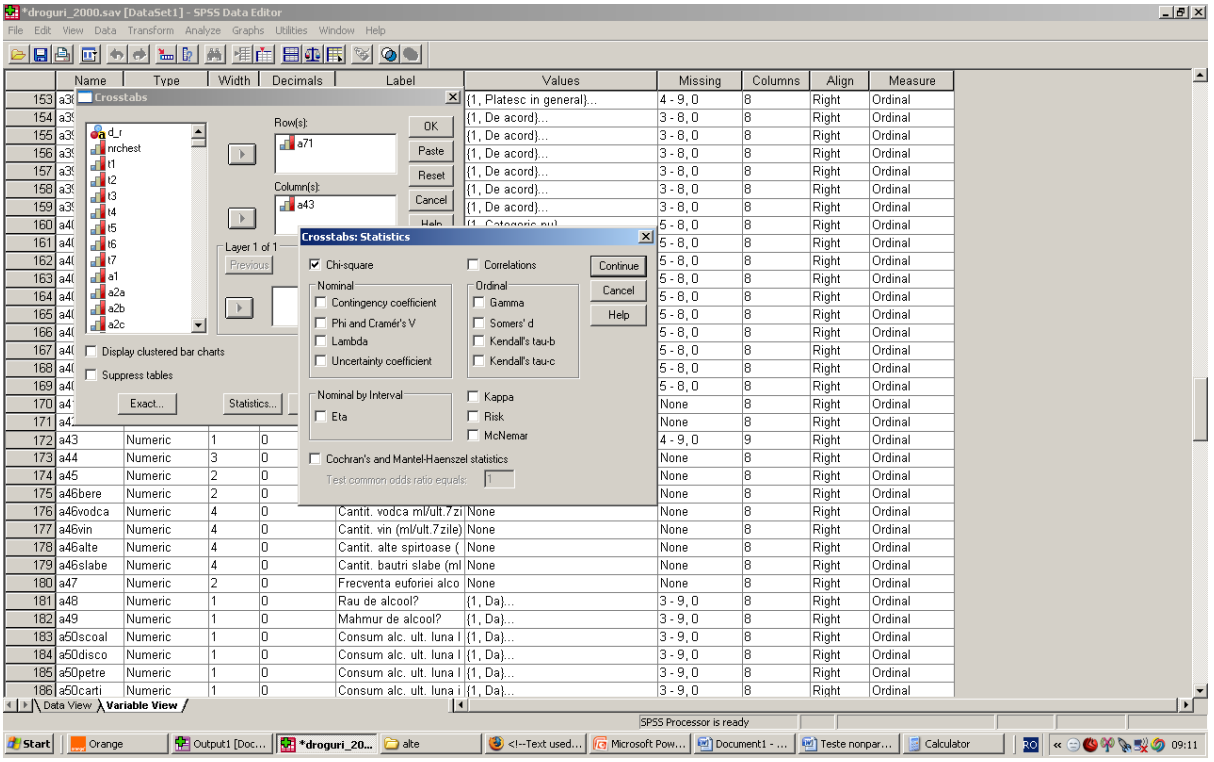
Unde:

T – frecvențele teoretice

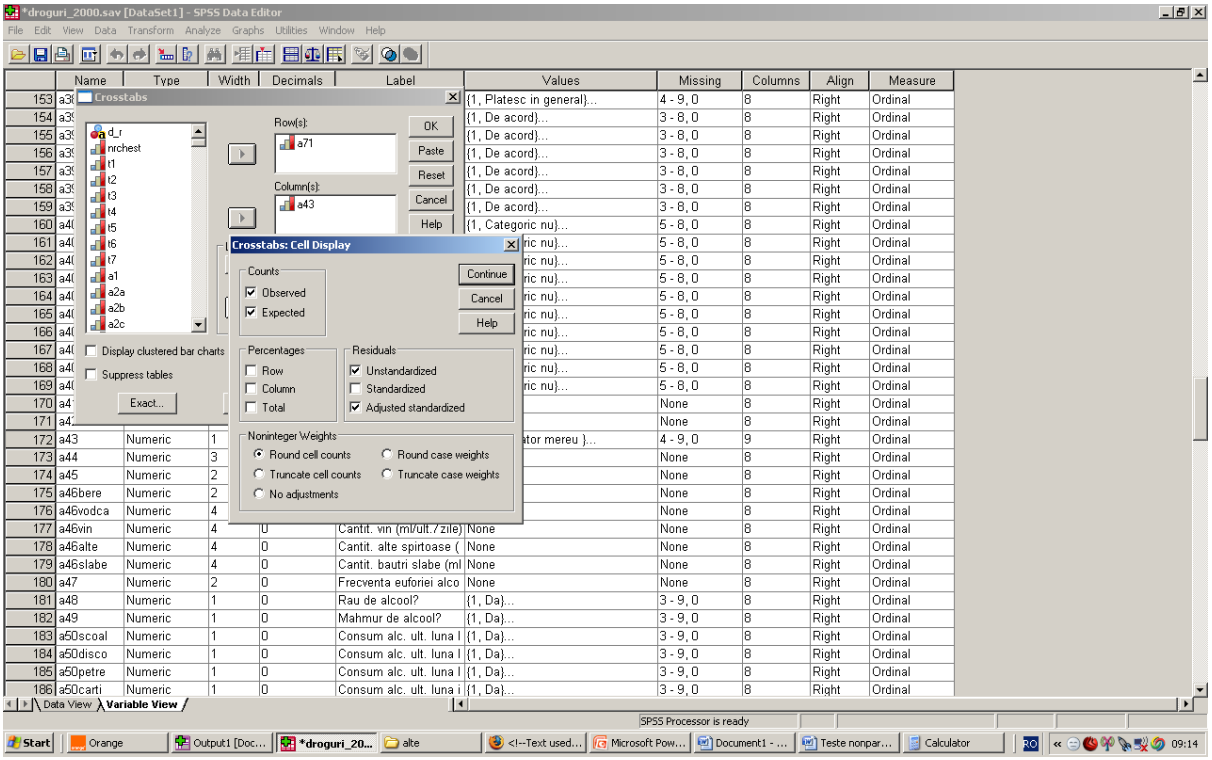
O – frecvențele observate

Semnificația statistică a lui Chi-pătrat va fi dată de mărimea testului de semnificație. Dacă acesta este mai mic de 0,05 respingem ipoteza nulă şi declarăm existența asocierii dintre cele două variabile.

Executarea testului chi-pătrat în SPSS este simplă. Din fereastra Crosstabs apăsați butonul Statistics iar dintre opțiunile activate selectați-o pe prima: Chi-square.



Apăsați Continue, ceea ce duce la închiderea ferestrei Crosstabs: Statistics şi apăsați butoul Cells din Fereastra Crosstabs. Din fereastra care se deschide activăm mai multe opțiuni: Observed şi Expected la Counts, Unstandardized şi Adjusted Standardized la Residuals. Rezultatele pe care aceste opțiuni ni le vor furniza nu sunt necesare în toate analizele de asociere. Le vom solicita în acest caz pentru a explica mai bine logica testului Chi-pătrat.



Apăsăm Continue apoi OK în fereastra de dialog Crosstabs. Rezultatul scris în output este prezentat în continuare (în format .doc).

[DataSet1] C:\spatiul de lucru\voyager\paginawebscurta\date\droguri_2000.sav

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Sex * Starea cu fumatul	955	95,5%	45	4,5%	1000	100,0%

Sex * Starea cu fumatul Crosstabulation

Frecvențe interioare observate.

Frecvențe interioare teoretice calculate după procedura expusă mai sus

Diferența dintre frecvențele teoretice şi cele observate

Reziduurile standardizate ajustate.

			Starea cu fumatul			Total
			Nefumator mereu	Am incercat doar	Sunt/am fost fumator	Nefumator mereu
		Count	121	171	134	42
		Expected Count	135,6	178,9	111,5	
			426, Residual	-14,6		-7,9
		22,5				
		Adjusted Residual	-2,0	-1,0	3,3	9
	Femeie	Count	183	230	116	52
		Expected Count	168,4	222,1	138,5	5
		, Residual	14,6	7,9	-22,5	
		Adjusted Residual	2,0	1,0	-3,3	
Total		Count	304	401	250	955
		Expected Count	304,0	401,0	250,0	955,0

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	11,648(a)	2	,003
Likelihood Ratio	11,613	2	,003
Linear-by-Linear Association	10,094	1	,001
N of Valid Cases	955		

Testele de semnificație ale primelor două măsuri ale asocierii au valori mai mici de 0,05. Se respinge, prin urmare, ipoteza nulă şi constatăm faptul că între sexul adolescentului şi fumat

i ă i

a 0 cells (,0%) have expected count less than 5. The minimum expected count is 111,52.

Toate detaliile de mai sus sunt clare, mai puțin reziduurile standardizate ajustate. Acestea sunt rezultatul transformării reziduurilor astfel încât să fie comparabile şi să aibă ca unitate de măsură abaterea standard a distribuției reziduurilor. Valoarea absolută a reziduului ajustat standardizat indică căsuțele care au abateri semnificative de la frecvențele teoretice (când reziduul ajustat e mai mare de 2) iar semnul indică direcția abaterii. În cazul căsuței folosite de mine ca referință, reziduul ajustat standardizat este -2, ceea ce indică o abatere negativă semnificativă. În cazul căsuței discutate mai sus (băiat fumător) abaterea este chiar mai mare dar pozitivă (3,3).

Testele de asociere (în rezultat avem mai multe, primul este chi-pătrat iar al doilea este un test similar) arată că între cele două variabile există asociere: fumatul depinde de sex, băieții având un risc mult mai mare decât fetele de a se declara fumătoare.

Reguli pentru realizarea unui test chi-pătrat corect

1. Se lucrează doar cu frecvențe, nu cu procente!
2. Toate căsuțele tabelului trebuie să fie nevide.
3. Procentul căsuțelor cu frecvențe teoretice mai mici de 5 nu poate depăşi 20% din totalul căsuțelor.

În cazul în care regulile 2 şi/sau 3 nu sunt realizate trebuie modificată distribuția variabilei, prin proceduri de recodificare, de obicei.

Alte măsuri asociate testului chi-pătrat

Testul exact al lui Fisher (Fisher's exact test)

Testul exact Fisher este raportat doar la tabele de 2 x 2. El poate fi folosit în loc de chi-pătrat când una sau mai multe din frecvențele teoretice are o valoare mai mică de 5.

Linear by linear measure of association

Această măsură a asocierii este adecvată numai dacă atât variabila de pe rând cât şi cea de pe coloană sunt cel puțin ordinale.

Alte măsuri ale asocierii

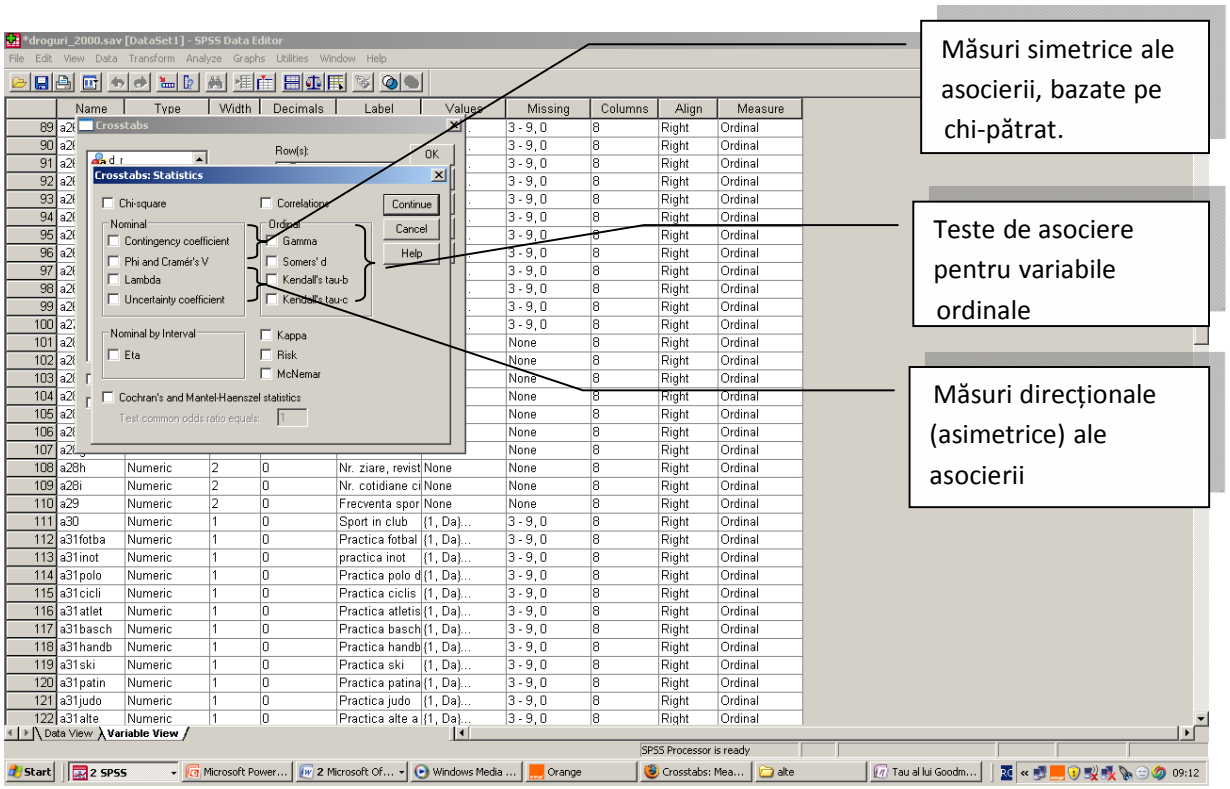
Conf. Dr. Adrian Hatos

Coeficientul chi-pătrat are mai multe dezavantaje:

- Nu indică intensitatea asocierii
- Nu are valori comparabile, mărimea testului fiind determinată de numărul de cazuri şi de numărul de grade de libertate
- Nu indică direcția asocierii – aspect interesant pentru variabile ordinale

Ca alternative standardizate la testul chi-pătrat au fost propuse mai multe categorii de teste a căror aplicare depinde de tipul de variabile implicat (nominal, ordinal etc.) şi de forma tabelului.

Testele statistice pot fi, de asemenea, clasificate în direcționale şi nedirecționale (simetrice). Cele direcționale (asimetrice) sunt cele în care se presupune o relație de cauzalitate între variabilele implicate.



Măsuri ale reducerii proporționale ale erorii de predicție. Coeficientul lambda

Coeficientul Chi-pătrat nu ne oferă informații despre intensitatea asocierii dintre cele două variabile.

Coeficientul λ (lambda) este un coeficient de reducere proporțională a erorii, la fel ca şi coeficienții de incertitudine şi tau al lui Goodman şi Kruskall.

Semnificația coeficientului- reducerea proporțională a erorii în predicția valorilor variabilei efect prin cunoaşterea valorilor variabilei cauză. Pentru a înțelege modul de calcul al coeficientului vom încerca să vedem care este relația dintre gen şi practicarea fotbalului în adolescență. Tabelul de asociere simplu al celor două variabile este prezentat mai jos.

Sex * Practica fotbal Crosstabulation

Count

		Practica fotbal		Total
		Da	Nu	Da
Sex	Barbat	298	101	399
	Femeie	69	372	441
Total		367	473	840

Care este eroarea de predicție a valorii variabilei efect dacă ştim doar frecvențele marginale? În acest caz, eroarea cea mai mică ar fi dacă am presupune că nu se practică fotbalul (valoarea cu frecvența cea mai mare: 473). Eroarea de predicție (e1) se va manifesta în restul cazurilor: 367.

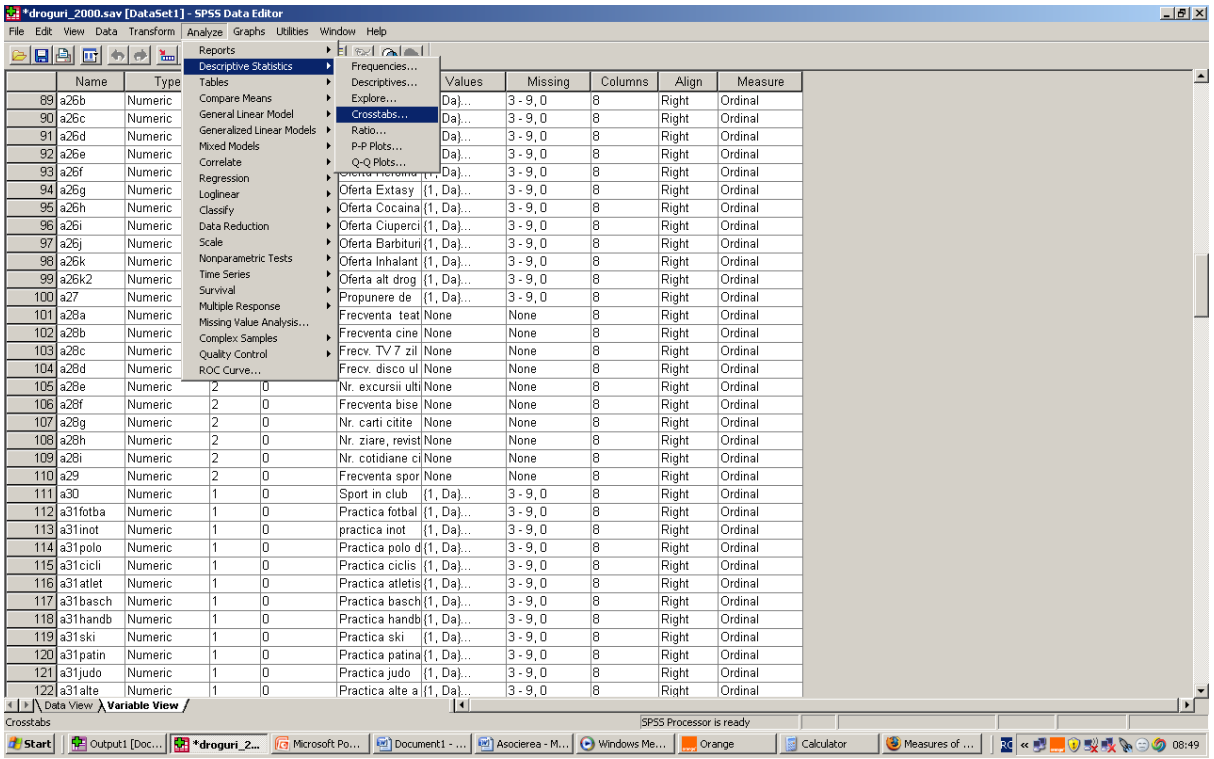
Cunoaşterea genului reduce puternic eroarea: dacă ştim că subiectul este bărbat vom presupune că practică fotbal (valoarea cu frecvența cea mai mare), greşind în doar 101 de cazuri iar dacă ştim că este femeie vom proceda analog, prezicând că nu practică fotbal şi greşind doar în 69 de cazuri. Cunoscând genul subiectului, numărul de erori de predicție se reduce la 69+101, adică 170 (e2).

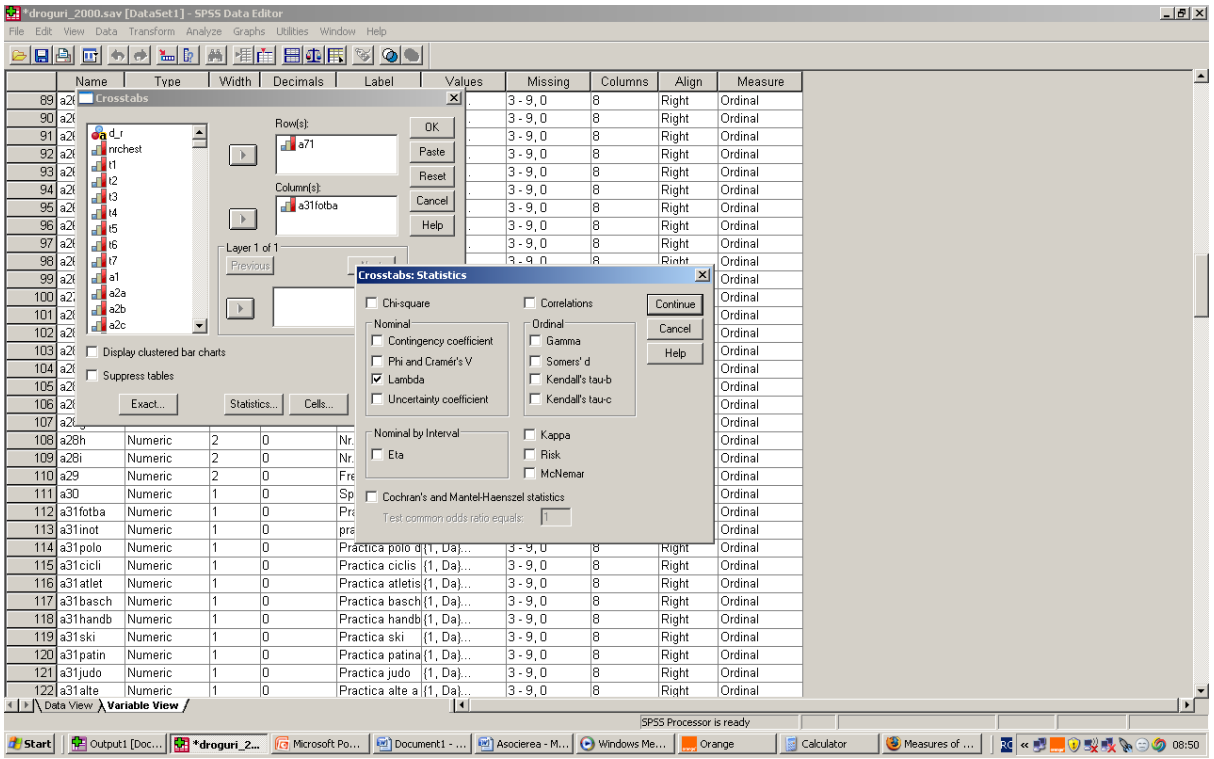
λ=(e1-e2)/e1 =0,537

Acelaşi lucru se poate calcula şi pentru situația în care încercăm să prezicem sexul persoanei cunoscând valorile caracteristicii practicare a fotbalului. În acest caz **λ=**0,573.

O măsură agregată a celor doi coeficienți denumiți asimetrici se poate calcula făcând media acestor doi coeficienți: 0,555.

Comanda pentru calcularea coeficientului lambda se aplică din acelaşi meniu ca şi în cazul coeficientului chi-pătrat.


*droguri_2000.sav [DataSet1] - SPSS Data Editor
File Edit View Data Transform Analyze Graphs Utilities Window Help
Reports
Descriptive Statistics
Tables
Compare Means
General Linear Model
Generalized Linear Models
Mixed Models
Correlate
Regression
Loglinear
Classify
Data Reduction
Scale
Nonparametric Tests
Time Series
Survival
Multiple Response
Missing Value Analysis...
Complex Samples
Quality Control
ROC Curve...
Frequencies...
Descriptives...
Explore...
Crosstabs...
Ratio...
P-P Plots...
Q-Q Plots...
Name Type Width Decimals Label Values Missing Columns Align Measure
89 a26b Numeric Oferta Extasy {1, Da}... 3 - 9, 0 8 Right Ordinal
90 a26c Numeric Oferta Cocaina {1, Da}... 3 - 9, 0 8 Right Ordinal
91 a26d Numeric Oferta Ciuperci {1, Da}... 3 - 9, 0 8 Right Ordinal
92 a26e Numeric Oferta Barbituri {1, Da}... 3 - 9, 0 8 Right Ordinal
93 a26f Numeric Oferta Inhalant {1, Da}... 3 - 9, 0 8 Right Ordinal
94 a26g Numeric Oferta alt drog {1, Da}... 3 - 9, 0 8 Right Ordinal
95 a26h Numeric Propunere de {1, Da}... 3 - 9, 0 8 Right Ordinal
96 a26i Numeric Frecventa teat None None 8 Right Ordinal
97 a26j Numeric Frecventa cine None None 8 Right Ordinal
98 a26k Numeric Frecv. TV 7 zil None None 8 Right Ordinal
99 a26k2 Numeric Frecv. disco ul None None 8 Right Ordinal
100 a27 Numeric Nr. excursii ulti None None 8 Right Ordinal
101 a28a Numeric Frecventa bise None None 8 Right Ordinal
102 a28b Numeric Nr. carti citite None None 8 Right Ordinal
103 a28c Numeric Nr. ziare, revist None None 8 Right Ordinal
104 a28d Numeric Nr. cotidiane ci None None 8 Right Ordinal
105 a28e Numeric 2 0 Frecventa spor None None 8 Right Ordinal
106 a28f Numeric 2 0 Sport in club {1, Da}... 3 - 9, 0 8 Right Ordinal
107 a28g Numeric 2 0 Practica fotbal {1, Da}... 3 - 9, 0 8 Right Ordinal
108 a28h Numeric 2 0 practica inot {1, Da}... 3 - 9, 0 8 Right Ordinal
109 a28i Numeric 2 0 Practica polo d {1, Da}... 3 - 9, 0 8 Right Ordinal
110 a29 Numeric 2 0 Practica ciclis {1, Da}... 3 - 9, 0 8 Right Ordinal
111 a30 Numeric 1 0 Practica atletis {1, Da}... 3 - 9, 0 8 Right Ordinal
112 a31fotba Numeric 1 0 Practica basch {1, Da}... 3 - 9, 0 8 Right Ordinal
113 a31inot Numeric 1 0 Practica handb {1, Da}... 3 - 9, 0 8 Right Ordinal
114 a31polo Numeric 1 0 Practica ski {1, Da}... 3 - 9, 0 8 Right Ordinal
115 a31cicli Numeric 1 0 Practica patina {1, Da}... 3 - 9, 0 8 Right Ordinal
116 a31atlet Numeric 1 0 Practica judo {1, Da}... 3 - 9, 0 8 Right Ordinal
117 a31basch Numeric 1 0 Practica alte a {1, Da}... 3 - 9, 0 8 Right Ordinal
118 a31handb Numeric 1 0 Practica alte a {1, Da}... 3 - 9, 0 8 Right Ordinal
119 a31ski Numeric 1 0 Practica alte a {1, Da}... 3 - 9, 0 8 Right Ordinal
120 a31patin Numeric 1 0 Practica alte a {1, Da}... 3 - 9, 0 8 Right Ordinal
121 a31judo Numeric 1 0 Practica alte a {1, Da}... 3 - 9, 0 8 Right Ordinal
122 a31alte Numeric 1 0 Practica alte a {1, Da}... 3 - 9, 0 8 Right Ordinal
Data View Variable View
Crosstabs
SPSS Processor is ready
Start Output1 [Doc... *droguri_2... Microsoft Po... Document1 - ... Asocierea - M... Windows Me... Orange Calculator Measures of ... 08:49


*droguri_2000.sav [DataSet1] - SPSS Data Editor
Crosstabs
d_r
nrchest
t1
t2
t3
t4
t5
t6
t7
a1
a2a
a2b
a2c
Row(s):
a71
Column(s):
a31fotba
Layer 1 of 1
Previous
Display clustered bar charts
Suppress tables
Exact...
Statistics...
Cells...
OK
Paste
Reset
Cancel
Help
Crosstabs: Statistics
Chi-square
Correlations
Nominal
Contingency coefficient
Phi and Cramér's V
Lambda
Uncertainty coefficient
Ordinal
Gamma
Somers' d
Kendall's tau-b
Kendall's tau-c
Nominal by Interval
Eta
Kappa
Risk
McNemar
Cochran's and Mantel-Haenszel statistics
Test common odds ratio equals:
Continue
Cancel
Help
Data View Variable View
SPSS Processor is ready
Start Output1 [Doc... *droguri_2... Microsoft Po... Document1 - ... Asocierea - M... Windows Me... Orange Calculator Measures of ... 08:50

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Sex * Practica fotbal	840	84,0%	160	16,0%	1000	100,0%

Sex * Practica fotbal Crosstabulation

Count

		Practica fotbal		Total
		Da	Nu	Da
Sex	Barbat	298	101	399
	Femeie	69	372	441
Total		367	473	840

Directional Measures

			Value	Asymp. Std. Error(a)	Approx. T(b)	Approx. Sig.
Nominal by Nominal	Lambda	Symmetric	,556	,033	12,584	,000
		Sex Dependent	,574	,031	13,122	,000
		Practica fotbal Dependent	,537	,037	10,488	,000
	Goodman and Kruskal tau	Sex Dependent	,353	,033		,000(c)
		Practica fotbal Dependent	,353	,033		,000(c)

a Not assuming the null hypothesis.

b Using the asymptotic standard error assuming the null hypothesis.

c Based on chi-square approximation

Mărimea testului ne spune că putem înjumătăți eroarea în predicția valorilor unei variabile cunoscând valorile celeilalte variabile dar nu ne indică dacă rezultatul este semnificativ din punct de vedere statistic. Testul de semnificație, mai mic de 0,05 ne edifică din acest punct de vedere: se respinge ipoteza nulă (potrivit căreia λ=0) şi constatăm că reducerea erorii de predicție este semnificativă.

Tau al lui Goodman şi Kruskall este similar lui lambda. El poate fi interpretat ca exprimând scăderea relativă în proporția predicțiilor incorecte atunci când trecem de la prezicerea categoriilor de pe linii bazându-ne doar pe probabilitățile marginale (ca şi la lambda) la prezicerea categoriilor de pe linii pe baza proporțiilor condiționale atât ale liniilor cât şi ale coloanelor. Un coeficient similar este cel de incertitudine.

Măsuri ale asocierii bazate pe chi-pătrat

Dezavantajele lui chi-pătrat pot fi depășite folosind câteva măsuri de asociere care pornesc de la valoarea lui chi-pătrat. Acestea sunt phi, V al lui Cramer și coeficientul de contingență. Toate au valori care tind să varieze între 0 și 1 ceea ce face rezultatele comparabile.

Phi depinde de tăria relației dintre variabile și de mărimea eșantionului. Este calculat prin extragerea rădăcinii pătrate din rezultatul împărțirii lui chi-pătrat pa numărul de cazuri. Pentru tabele cu mai multe linii și coloane, phi poate depăși valoarea 1.

V al lui Cramer este cel mai popular test dintre cele bazate pe chi-pătrat deoarece variază doar între 0 și 1. V este calculat astfel:

$$V = \sqrt{x^2 / N(k-1)}$$

Unde N= numărul de cazuri iar k = cel mai mic dintre numărul de linii și numărul de coloane. Pentru tabele 2X2 V=phi.

Coeficientul de contingență

CC este o altă cale de a corecta dependența mărimii lui chi-pătrat de mărimea eșantionului.

Formula CC este:

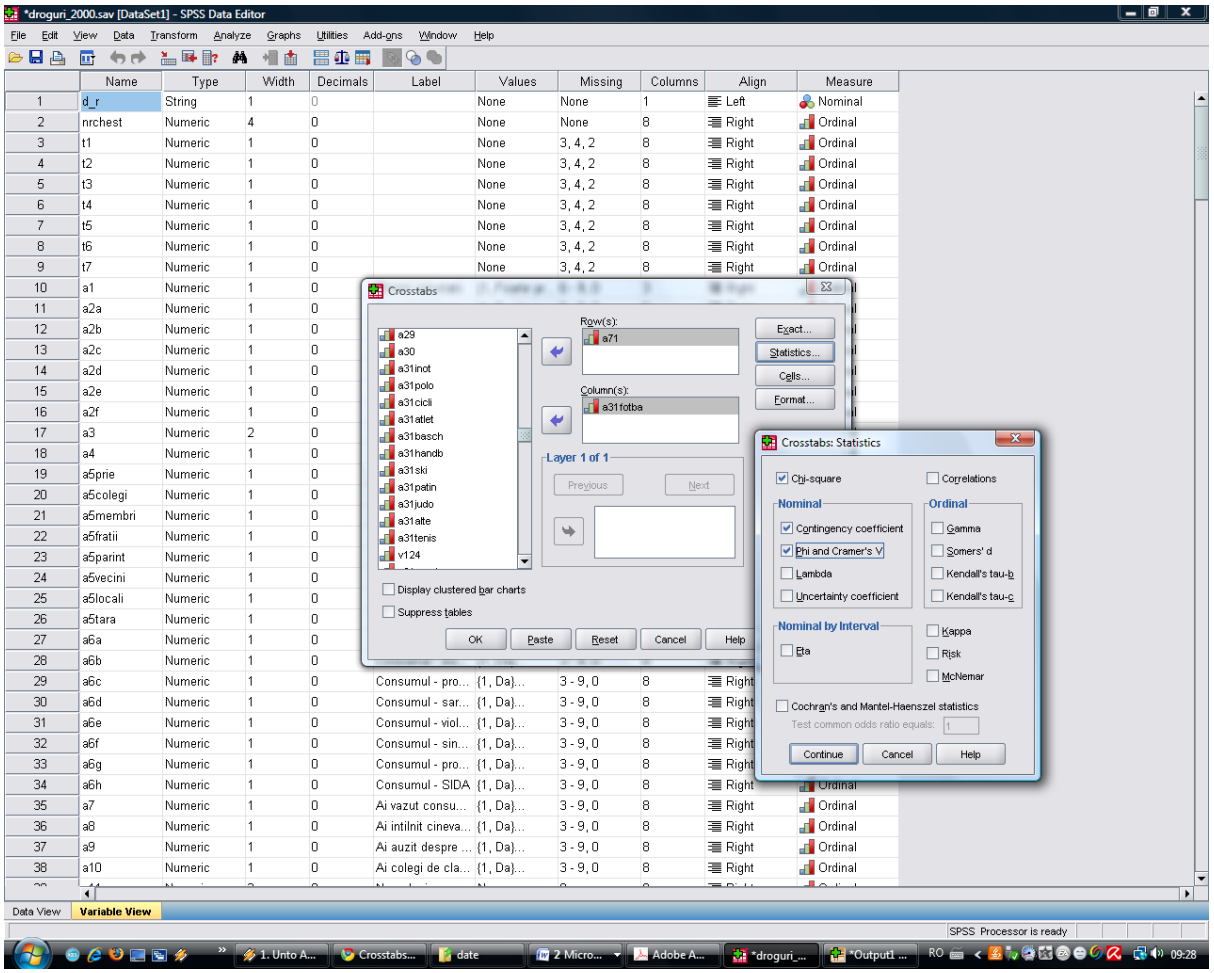
$$CC = \sqrt{x^2 / (x^2 + N)}$$

Unde N = numărul de cazuri.

Problema coeficientului de contingență este că valoarea sa maximă depinde de mărimea tabelului. Astfel, pentru un tabel 2X2 valoarea maximă este de 0,707 iar pentru un tabel 4X4 este de 0,87. Astfel, este imposibil să se compare coeficienți de contingență calculați pe tabel de mărimi diferite. Se recomandă, prin urmare, utilizare lui al lui Cramer.

Toți trei cei trei coeficienți sunt simetrici.

Pentru calcularea acestor indici se selectează opțiunile corespunzătoare din fereastra de dialog Crosstabs: Statistics. Apoi Continue și OK.



Rezultatul, inserat din format html.

[DataSet1] D:\lucru\paginawebscurta\date\droguri_2000.sav

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Sex * Practica fotbal	840	84,0%	160	16,0%	1000	100,0%

Sex * Practica fotbal Crosstabulation

Count				
		Practica fotbal		
		Da	Nu	Total
Sex	Barbat	298	101	399
	Femeie	69	372	441
	Total	367	473	840

Chi-Square Tests					
	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	2,968E2	1	,000		
Continuity Correction[b]	294,404	1	,000		
Likelihood Ratio	317,032	1	,000		
Fisher's Exact Test				,000	,000
Linear-by-Linear Association	296,446	1	,000		
N of Valid Cases	840				
a. 0 cells (,0%) have expected count less than 5. The minimum expected count is 174,33.					
b. Computed only for a 2x2 table					

Symmetric Measures			
		Value	Approx. Sig.
Nominal by Nominal	Phi	,594	,000
	Cramer's V	,594	,000
	Contingency Coefficient	,511	,000
	N of Valid Cases	840	

Cei trei coeficienți și testele de semnificație sunt în ultimul tabel. Testele de semnificație sunt, de fapt, preluate de la chi-pătrat. Pe baza acestora stabilim faptul că asocierea este semnificativă.

Pentru a evidenția valoarea comparabilității coeficienților statistici, am calculat măsurile asocierii genului și cu practicarea baschetului. Supoziția mea este că asocierea va fi mai slabă decât în cazul fotbalului, baschetul fiind mai puțin specific genului masculin.

Crosstab				
Count				
		Practica baschet		
		Da	Nu	Total
Sex	Barbat	178	178	356
	Femeie	175	284	459
	Total	353	462	815

Symmetric Measures			
		Value	Approx. Sig.
Nominal by Nominal	Phi	,119	,001
	Cramer's V	,119	,001
	Contingency Coefficient	,118	,001

	N of Valid Cases	815	

Chiar dacă asocierea este din nou semnificativ din punct de vedere statistic, ea este, într-adevăr mai slabă. Valoarea lui V este de doar 0,119 față de 0,594. Putem spune, prin urmare, că practicarea baschetului este mai puțin specifică genului masculin decât practicarea fotbalului.

Măsuri ale asocierii pentru variabile ordinale

Problemă: există relație între frecvența declarată a problemelor cu somnul și incidența stărilor de tristețe și deprimare?

Dacă pentru variabile nominale coeficienții prezentați până acum sunt suficienți, în cazul variabilelor ordinale pot să fie socotiți nesatisfăcători deoarece în cazul unor astfel de variabile putem avea pretenția legitimă de a fi măsurat și sensul asocierii – în sensul covariației. Adică dacă valorile tind să concorde sau să fie discordante.

SPSS calculează mai mulți coeficienți pentru variabile ordinale care toate pleacă de la compararea numărului de perechi concordante cu cel al numărului de perechi discordante de cazuri.

Coeficientul gamma, de pildă , are următoarea formulă:

$$\gamma = (P - Q) / (P + Q)$$

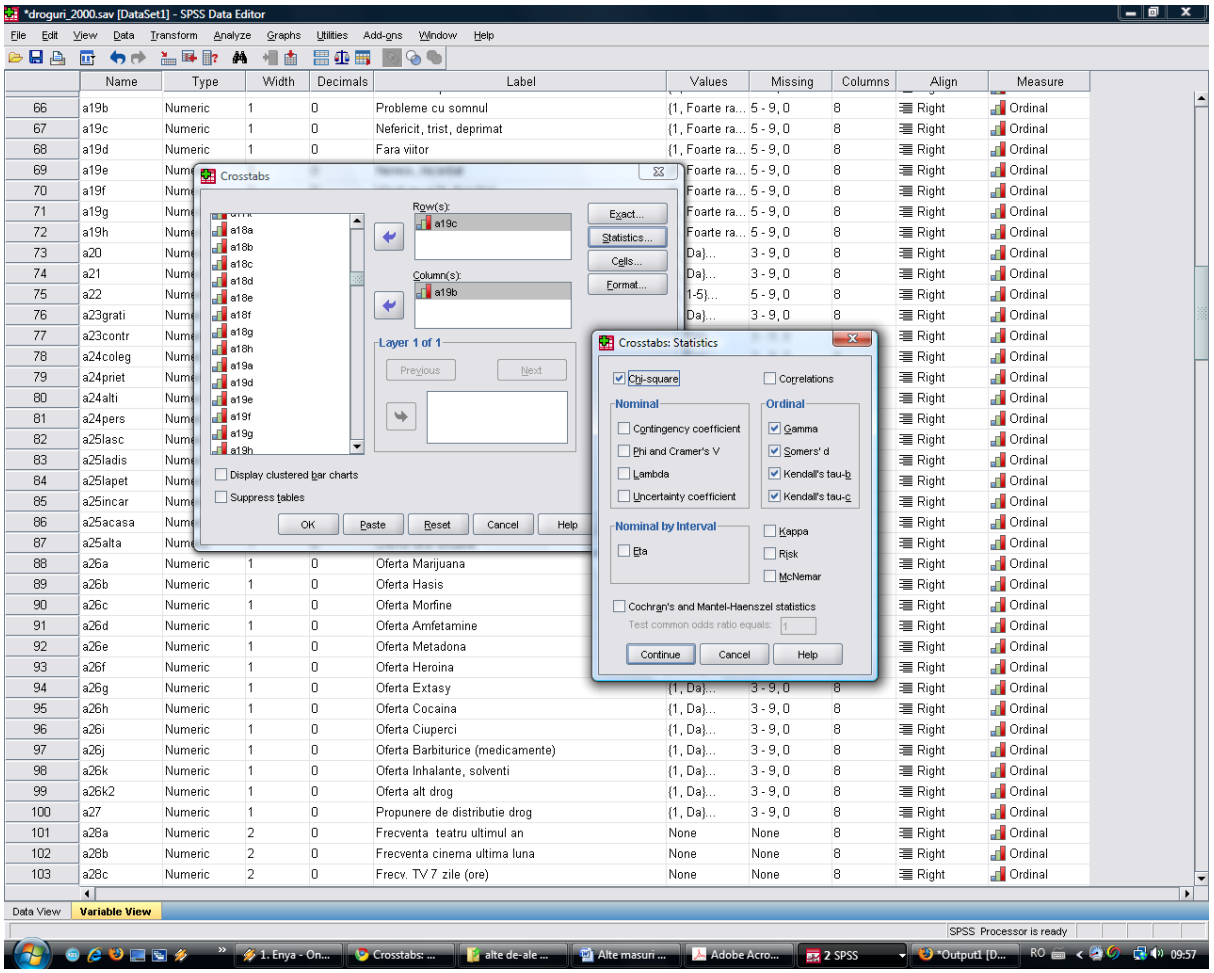
Unde P = numărul de perechi concordante iar Q numărul de perechi discordante. Pentru detalii referitoare la calcularea lui gamma puteți consulta pagina de web:
http://web.uccs.edu/lbecker/SPSS/ctabs2.htm#5B. Gamma, g

Tau-C al lui Kendall și Tau b al lui Kendall sunt coeficienți similari lui Gamma care fac corecții pentru numărul de linii și coloane și folosesc toate cazurile.

În plus, SPSS calculează și un coeficient direcțional pentru variabile ordinale – **d al lui Somers**, care se folosește cu premisa că una dintre variabile este cauză iar alta efect.

Soluție în SPSS

Coeficienții ordinali pot fi selectați în fereastra Crosstabs: Statistics. Apoi Continue și OK în fereastra principală.



Rezultatul este mai jos:

[DataSet1] D:\lucru\paginawebscurta\date\droguri_2000.sav

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Nefericit, trist, deprimat * Probleme cu somnul	902	90,2%	98	9,8%	1000	100,0%

Nefericit, trist, deprimat * Probleme cu somnul Crosstabulation

Count

		Probleme cu somnul				
		Foarte rar	Rar	Des	Foarte des	Total
Nefericit, trist, deprimat	Foarte rar	138	47	14	4	203
	Rar	166	122	53	21	362
	Des	68	102	51	18	239

	Foarte des	26	26	21	25	98
	Total	398	297	139	68	902

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	1,299E2	9	,000
Likelihood Ratio	118,485	9	,000
Linear-by-Linear Association	96,736	1	,000
N of Valid Cases	902		

a. 0 cells (,0%) have expected count less than 5. The minimum expected count is 7,39.

Directional Measures

			Value	Asymp. Std. Error[a]	Approx. T[b]	Approx. Sig.
Ordinal by Ordinal	Somers' d	Symmetric	,286	,027	10,347	,000
		Nefericit, trist, deprimat Dependent	,294	,028	10,347	,000
		Probleme cu somnul Dependent	,278	,027	10,347	,000

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

Symmetric Measures

		Value	Asymp. Std. Error[a]	Approx. T[b]	Approx. Sig.
Ordinal by Ordinal	Kendall's tau-b	,286	,027	10,347	,000
	Kendall's tau-c	,262	,025	10,347	,000
	Gamma	,408	,037	10,347	,000
	N of Valid Cases	902			

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

Din tabelele de mai sus, pe baza testelor de semnificație, constatăm că toți coeficienții de asociere sunt semnificativi (sig<0,05).

Coeficienți de risc

Problemă: Cu cât este mai mare șansa ca un băiat să practice fotbalul decât șansa similară pentru o fată?

Dacă a doua variabilă ar fi desemnat ceva periculos, precum apariția unei boli sau un comportament de risc, putea folosi liniștit în locul conceptului neutru de șansă pe cel de risc. Să reluăm tabelul de asociere al celor două variabile:

Sex * Practica fotbal Crosstabulation

Count

		Practica fotbal		
		Da	Nu	Total
Sex	Barbat	298	101	399
	Femeie	69	372	441
	Total	367	473	840

Șansa ca un băiat să joace fotbal raportat la șansa de a nu juca fotbal este de 298/101=2,95

Șansa ca o fată să joace fotbal raportată la șansa de a nu juca este de 69/372=0,185.

Raportul de șanse (odds ratio) se calculează împărțind cele două șanse:

Odds ratio (băiat/față)=2,95/0,185=15,94

Coeficientul arată că băieții au o probabilitate mult mai mare de a juca fotbal decât băieții.

Pentru a reține ușor formula folosim exemplul teoretic de mai jos. Plecăm de la premisa unei asocieri de două dihotomii care au valorile 1 și 2, și, respectiv, a și b.

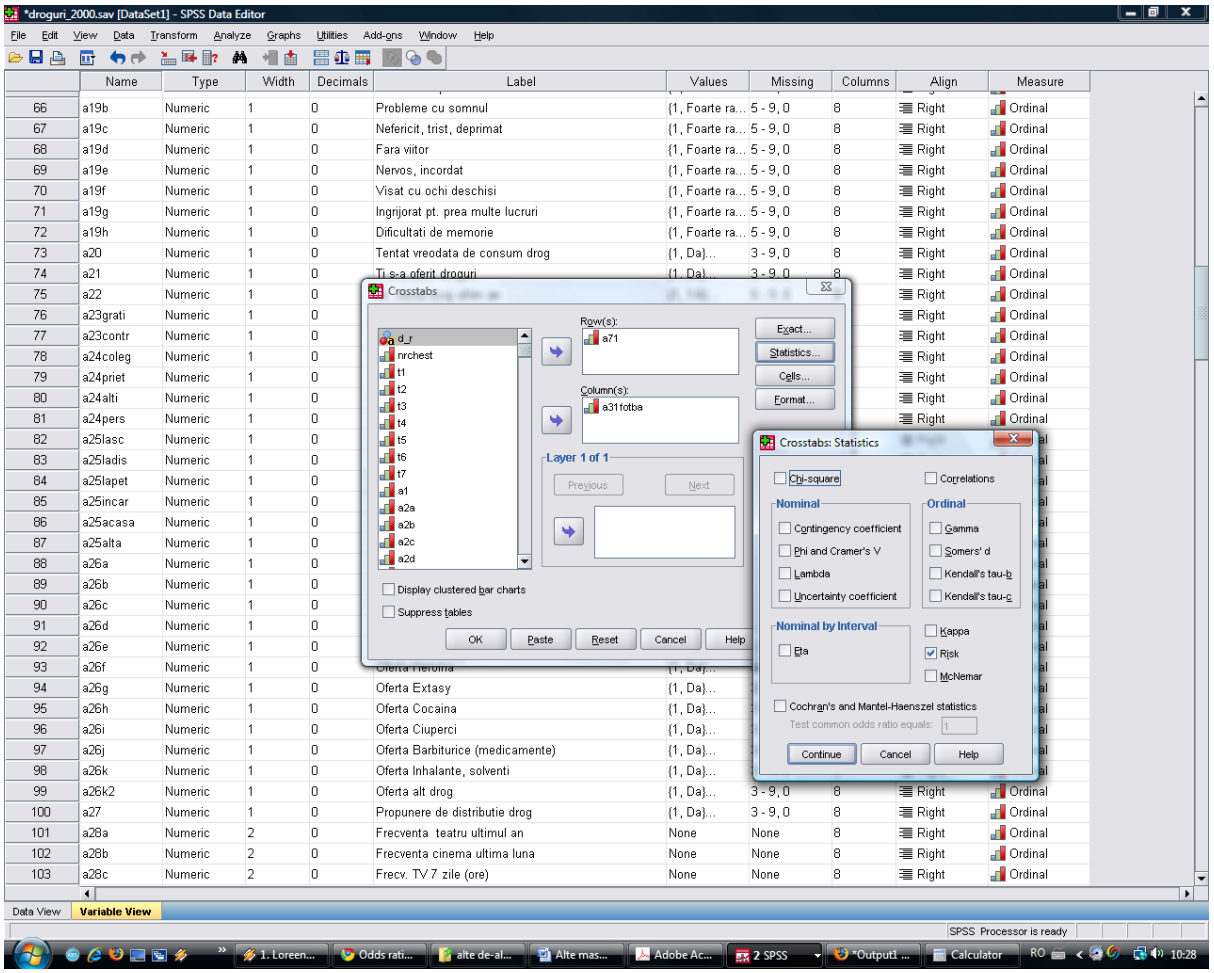
	a	b
1	1a	1b
2	2a	2b

Odds-ratio=1a2b/1b2a

Cu alte cuvinte, odds-ratio este rezultatul împărțirii produsului valorilor de pe diagonala principală (care, de multe ori semnifică potriviri în termeni de da/nu sau adevărat/fals) la produsul valorilor de pe diagonala secundară.

Soluție SPSS

Selectăm opțiunea Risk din fereastra Crosstabs: Statistics. Apoi Continue și OK.



Rezultatul exportat în html, mai jos.

Risk Estimate

	Value	95% Confidence Interval	
		Lower	Upper
Odds Ratio for Sex (Barbat / Femeie)	15,907	11,300	22,392
For cohort Practica fotbal = Da	4,773	3,815	5,973
For cohort Practica fotbal = Nu	,300	,252	,357
N of Valid Cases	840		

Observăm că rezultatul obținut este similar cu cel obținut de mine (diferența este produsă de rotunjirile operate de mine) chiar dacă calculele intermediare de șanse au fost altele.

Mai trebuie remarcat că valoarea lui odds ratio depinde de ordinea valorilor. Dacă, de pildă, în exemplul meu categoria Fată ar fi fost înaintea celei a băieților, rezultatul ar fi fost 0,062. În plus, odd-ratio variază neliniar de numărul de cazuri și de distribuția acestora.

Aceste defecte sunt înlăturate prin utilizarea valorii logaritmate a lui odds-ratio – log-odds.

De pildă:

ln15,907=2,76

ln0,062=-2,78[1]

Odds ratio este folosit frecvent pentru a exprima diferențe de probabilitate pentru orice dihotomii.

[1] Diferența se datorează din nou rotunjirilor. Cele două valori ar fi trebuit să fie identice în valoare absolută.

Compararea mediilor pe două grupuri independente (testul t)

Problemă: diferă numărul mediu de țigări fumate de băieți în ziua anterioară realizării anchetei față de măsura identică în cazul fetelor?

Atunci când testăm ipoteza nulă că mediile a două populații independente distribuite normal sunt egale folosim testul t.

Normalitatea distribuției celor două grupuri se testează cu teste de normalitate precum Shapiro-Wilk sau Kolmogorov-Smirnov. În cazul în care testele de normalitate dau rezultate negative se recomandă folosirea altor teste, precum testul U al lui Mann-Whitney.

Testul t are formula generala:

T=a-b/SE,

unde a și b sunt cele două mărimi care trebuie comparate iar SE este eroarea standard.

Avem de-a face, de fapt, cu mai multe teste care depind de două condiții: egalitatea de mărime a eșantioanelor și egalitatea variațelor.

Eșantioane de mărimi egale, variațe egale

În cazul în care mărimile populațiilor sunt similare iar variațele sunt similare, formula de calcul a lui t este următoarea:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{S_{X_1X_2} \cdot \sqrt{\frac{2}{n}}}$$

unde

$$S_{X_1X_2} = \sqrt{\frac{S_{X1}^2 + S_{X2}^2}{2}}.$$

S_{x1x2}= abaterea standard comună a celor două populații iar S_{x1} și S_{x2} sunt abaterile standard ale celor două eșantioane.

n=numărul de cazuri

Eșantioane de mărimi inegale, variațe egale

$$t = \frac{\bar{X}_1 - \bar{X}_2}{S_{X_1X_2} \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

unde

$$S_{X_1X_2} = \sqrt{\frac{(n_1 - 1)S_{X1}^2 + (n_2 - 1)S_{X2}^2}{n_1 + n_2 - 2}}.$$

Eșantioane de mărimi inegale și variante inegale

$$t = \frac{\overline{X}_1 - \overline{X}_2}{s_{\overline{X}_1 - \overline{X}_2}}$$

unde

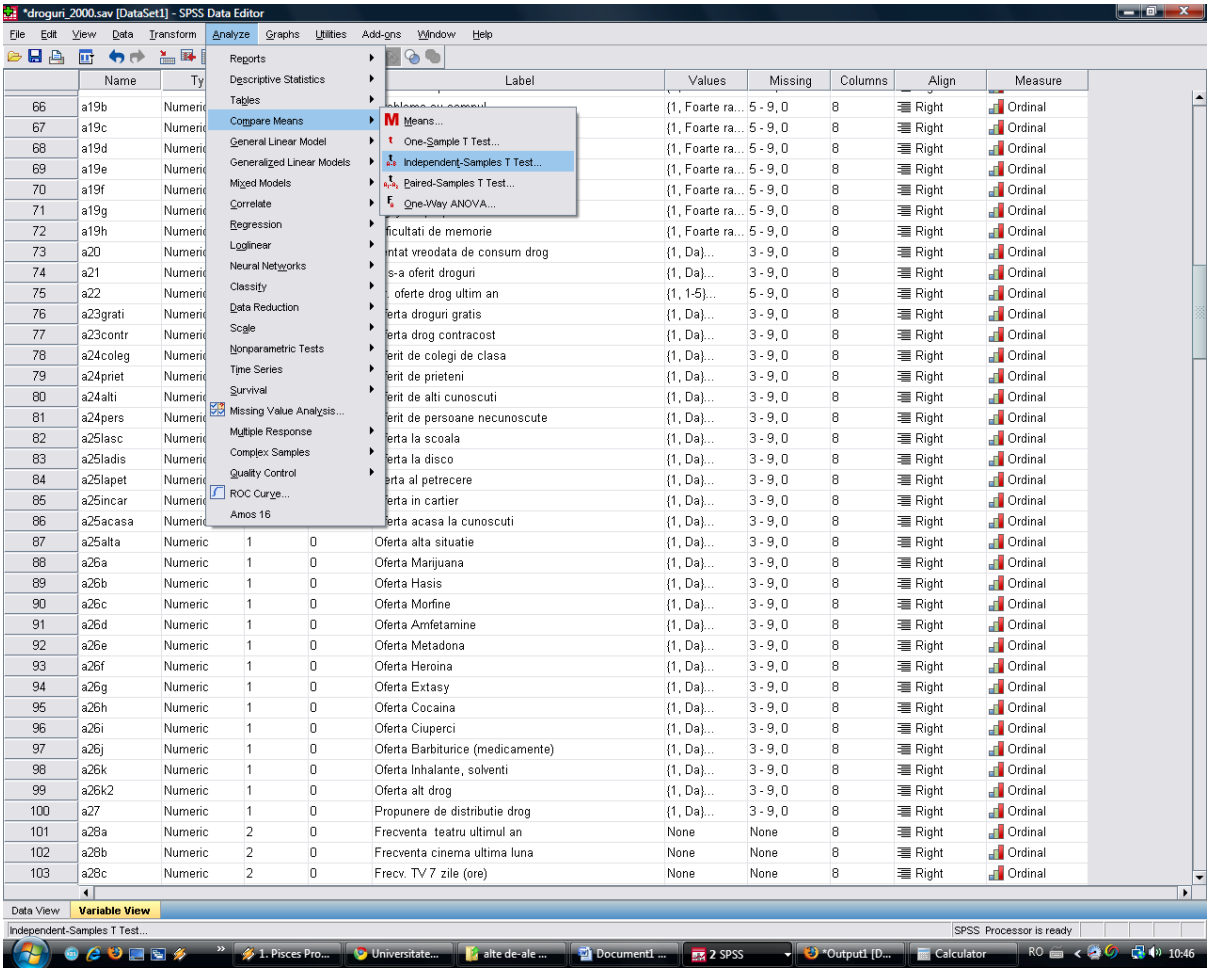
$$s_{\overline{X}_1 - \overline{X}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

S^2= un estimat nedistorsionat al variantei celor două eșantioane.

SPSS decide după algoritmi interni dacă mărimile celor două eșantioane diferă sau nu, în schimb decizia privind egalitatea variantelor aparține cercetătorului. În fapt, SPSS calculează testul t atât pentru situația de egalitate a variantelor cât și pentru situația alternativă rămânând la latitudinea cercetătorului rezultatul pe care îl va folosi, în funcție de rezultatul testului F de egalitate a variantelor.

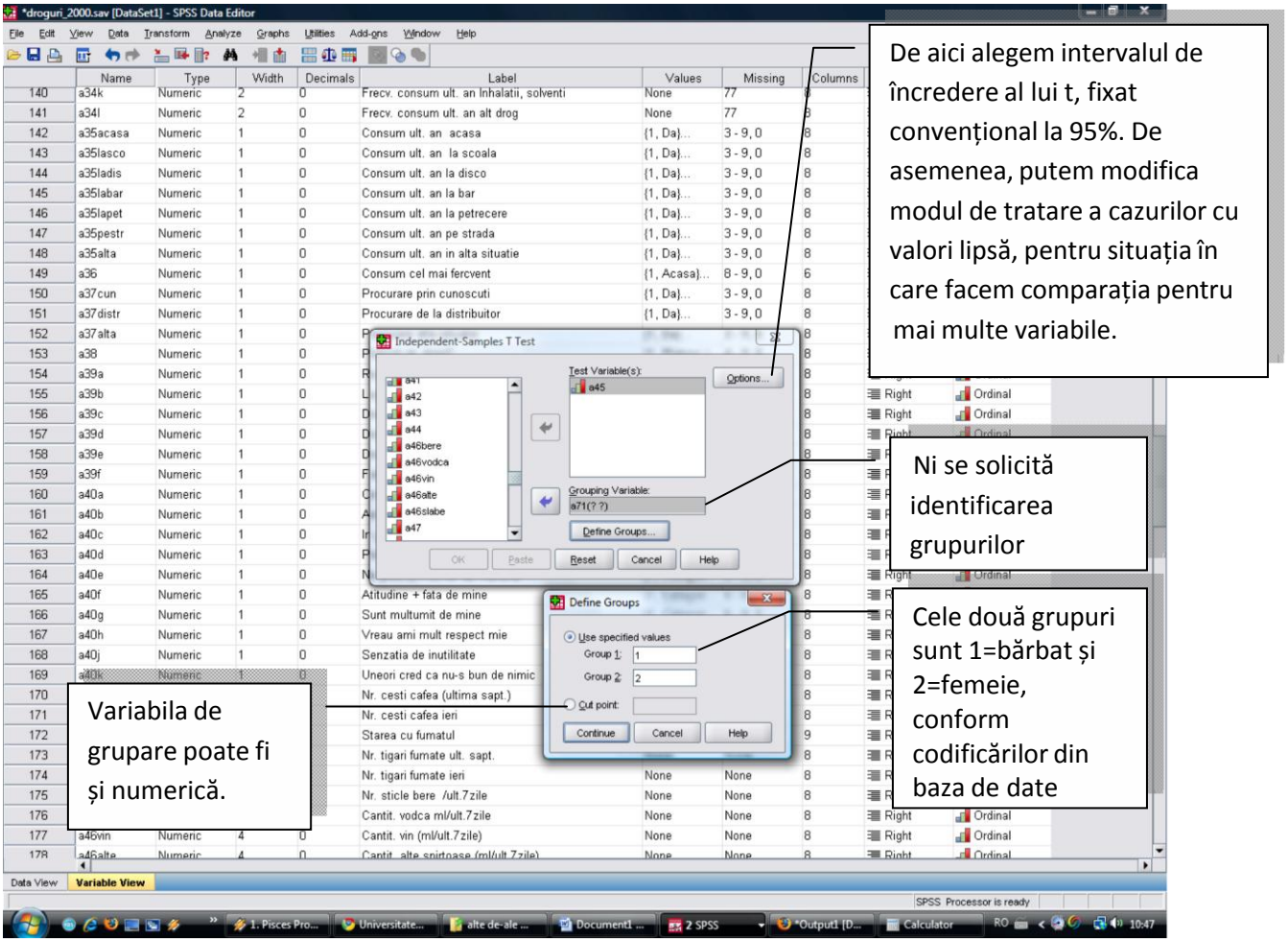
Soluție SPSS

Se alege opțiunea **Independent-samples T test** din submeniul Compare means.



Independent samples t-test compară prin testul t mediile calculate pe grupuri diferite (de exemplu, grup experimental vs grup de control). Dacă se compară măsurări realizate pe același grup (de exemplu, înainte-după) se folosește opțiunea **Paired-samples t test**.

Opțiunea **means** permite doar comparația simplă a mediilor**. One-sample t test** permite testarea de ipoteze referitoare la media unei variabile numerice. Este, prin urmare un test univariat.



După deschiderea ferestrei de dialog selectăm variabila care furnizează grupurile, pe care o trecem la categoria **test variable** și, apoi, variabila care generează grupurile, în cazul nostru genul (a71). Odată ce am stabilit că **Grouping variable** este a71 două semne de întrebare ne avertizează că trebuie să indicăm cele două grupuri. Pentru aceasta apăsăm butonul **Define groups**.

Grupurile pot fi marcate prin coduri sau ca punct de tăietură pe scala unei variabile categorice. Cu alte cuvinte, nu este obligatoriu ca variabila care generează grupurile să fie categorică.

Continue, apoi OK

Rezultatul exportat în html, mai jos.

[DataSet1] D:\lucru\paginawebscurta\date\droguri_2000.sav

Group Statistics

	Sex	N	Mean	Std. Deviation	Std. Error Mean
Nr. tigari fumate ieri	Barbat	382	3,14	7,765	,397
	Femeie	451	1,23	3,115	,147

Independent Samples Test

	Levene's Test for Equality of Variances	t-test for Equality of Means

		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Nr. tigari fumate ieri	Equal variances assumed	65,620	,000	4,808	831	,000	1,918	,399	1,135	2,701
	Equal variances not assumed			4,528	484,326	,000	1,918	,424	1,086	2,750

Primul tabel, descriptiv arată că, într-adevăr, între mediile celor două categorii avem o diferență apreciabilă: băieții fumează de aproape 2,6 ori mai multe țigări decât fetele de la liceu. Dar este această diferență semnificativă? Al doilea tabel ne lămurește prin testul t.

Acest tabel se citește în două etape:

1. Interpretarea testului F de egalitate a varianțelor și stabilirea rezultatului t pe care îl citim

Testul testează ipoteza nulă că varianțele sunt egale. Deoarece testul de semnificație al testului F este mai mic de 0,05, respingem ipoteza de egalitate a varianțelor și vom citi testul t de pe a doua linie corespunzătoare varianțelor inegale (Equal variances not assumed).

2. Interpretarea testului t

Mărimea testului este de 4,52. Semnificația testului este mai mică de 0,05, astfel că respingem ipoteza nulă, a egalității mediilor. Se poate spune că băieții de liceu fumează mai mult decât fetele.

Observație

Manualele de statistică folosesc denumiri diferite pentru acest gen de teste (testul t al lui Student, testul Z), formulele fiind diferite și în funcție de numărul de cazuri. SPSS știe să facă deosebirile necesare și să aplice formula corectă.

Coeficienţi de fidelitate

Conceptul de fidelitate al măsurării şi pentru scale

Folosind analiza de fidelitate puteți determina măsura în care itemii din chestionar sunt corelați între ei, puteți obține un indice general al repetabilității sau consistenței interne a scalei per ansamblu și puteți identifica itemi problematici care trebuie eliminați din scală.

Cerințe
Scala trebuie să fie aditivă (itemii să fie codați numeric) astfel încât fiecare să fie corelat liniar cu scala.

Proceduri corelate
Dimensionalitatea scalelor se explorează cu analiză factorială. Seturi omogene de variabile (care pot constitui o scală) pot fi identificate și cu ajutorul Clusterului Ierarhic (clasificare multicriterială).
Condiții

- Alpha Cronbach - bazat pe corelaţia medie dintre itemi

- Coeficienţii lui Guttman
    - L1. Coeficient intermediar folosit în calcularea celorlalţi 5
    - L2. Mai complex decât alpha al lui Cronbach, preferat de unii cercetători
    - L3. Echivalent cu alpha al lui Cronbach
    - L4. Fidelitatea split-half a lui Guttman
    - L5. Recomandată atunci când un item corelează puternic cu alţi itemi care întrei ei nu corelează puternic
    - L6. Recomandată atunci când corelaţiile dintre itemi sunt mici în comparaţie cu corelaţiile multiple pătratice.

În SPSS
Modele: Alpha, Split half, Guttman
Statistics: Descriptives for item, scale, scale if item deleted
Inter-item corrleations, Hotellings T-square

Problemă: verificați dimensionalitatea și fidelitatea itemilor de satisfacție cu viața a1-a2f din baza de date droguri2000.

Întrebările din problemă au următorul text:

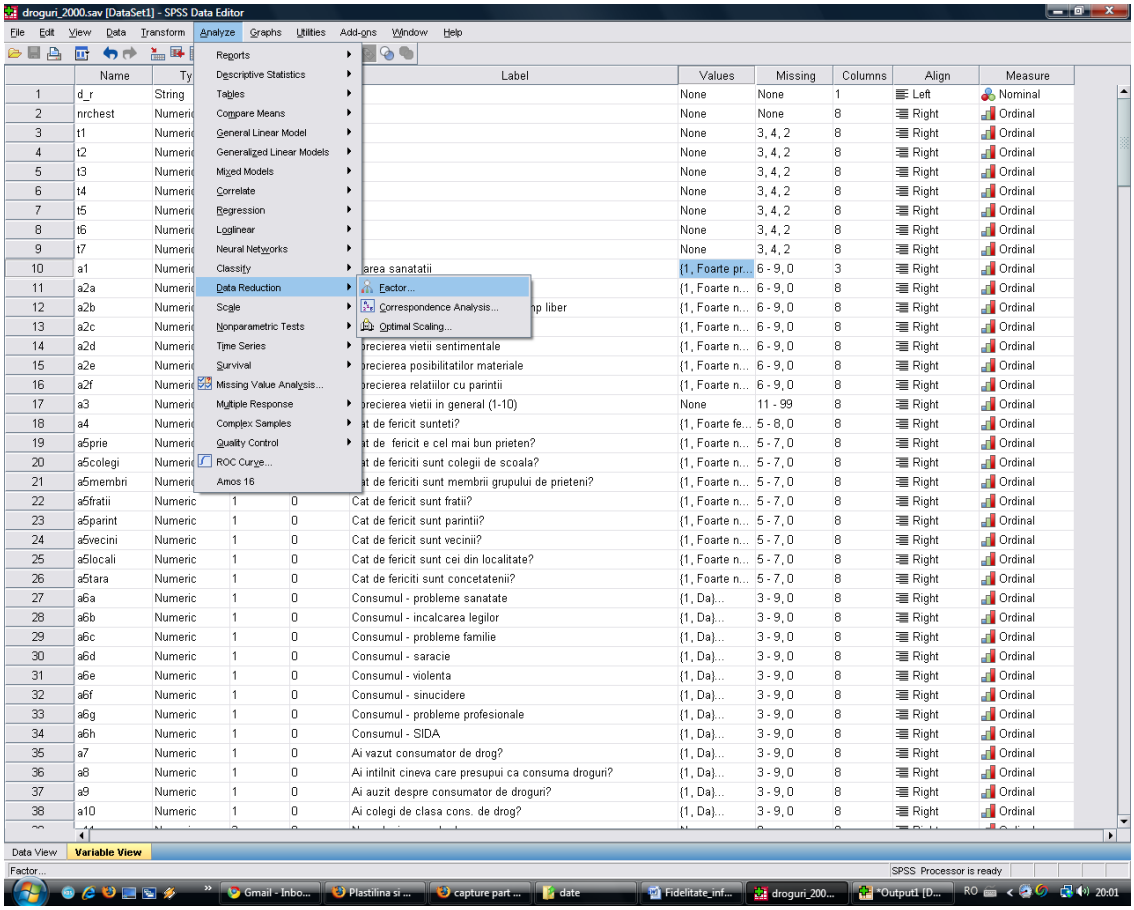
Cât de mulțumit ești de…
Starea sanatatii?
Rezultatele tale școlare?
Modul în care îți petreci timp liber?
Prietenii tăi?
Viața ta sentimentală?
Posibilitatile tale materiale?
Relatiile tale cu parintii?

Pentru fiecare item au fost oferite 5 variante de răspuns pe scară grafică de tip Likert. Formal, scala se pretează la construcția de indice prin adunarea simplă a scorurilor pe itemi. Totuși, pentru a construi un astfel de indice (al satisfacției cu viața) trebuie să fim siguri că scala este unidimensională și fidelă.

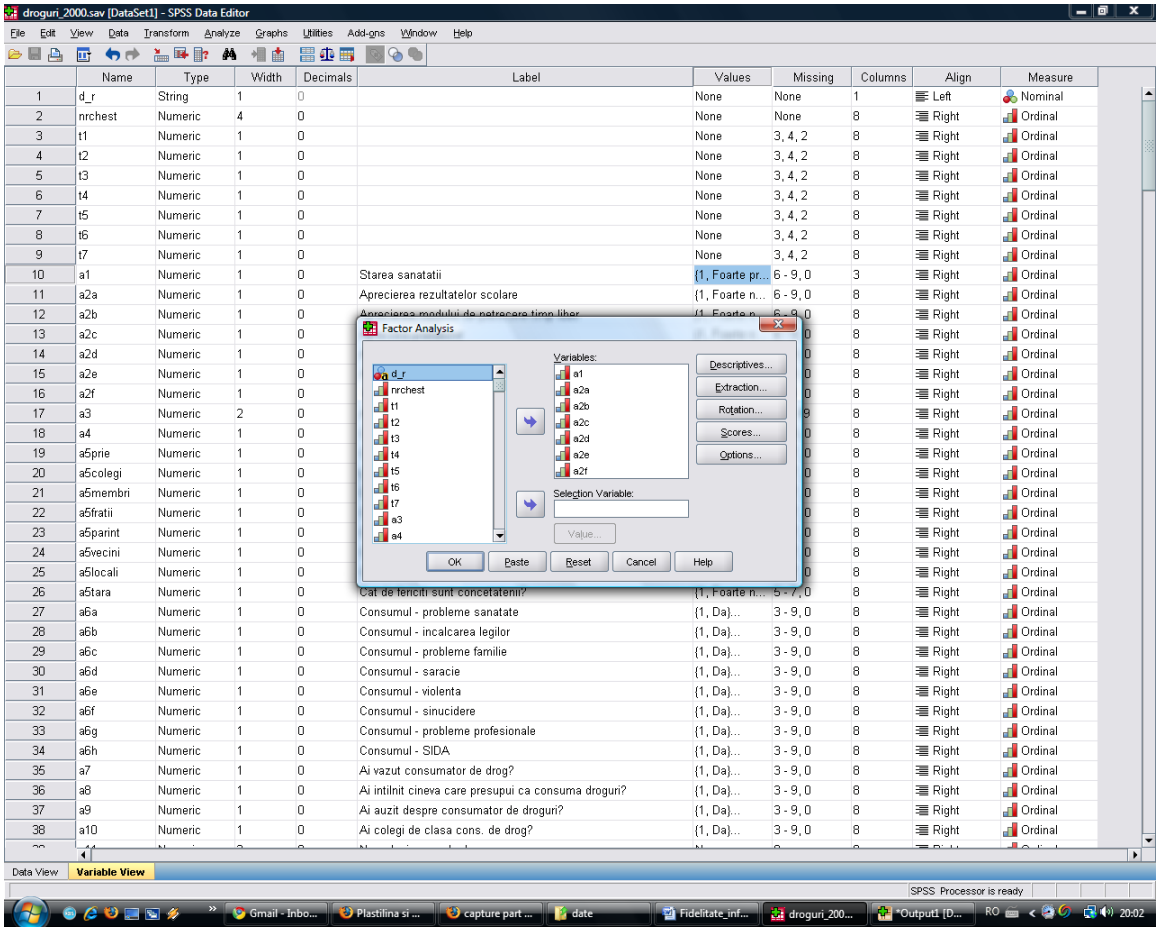
I.Analiza dimensionalității

Se realizează cu analiză factorială (metoda componentelor principale)

Verificați ca valorile lipsă (erori de introduce, valori aberante) ale variabilelor din baza de date să fie bine definite.
Alegeți analiza factorială din meniul *Analyze* din meniul principal.



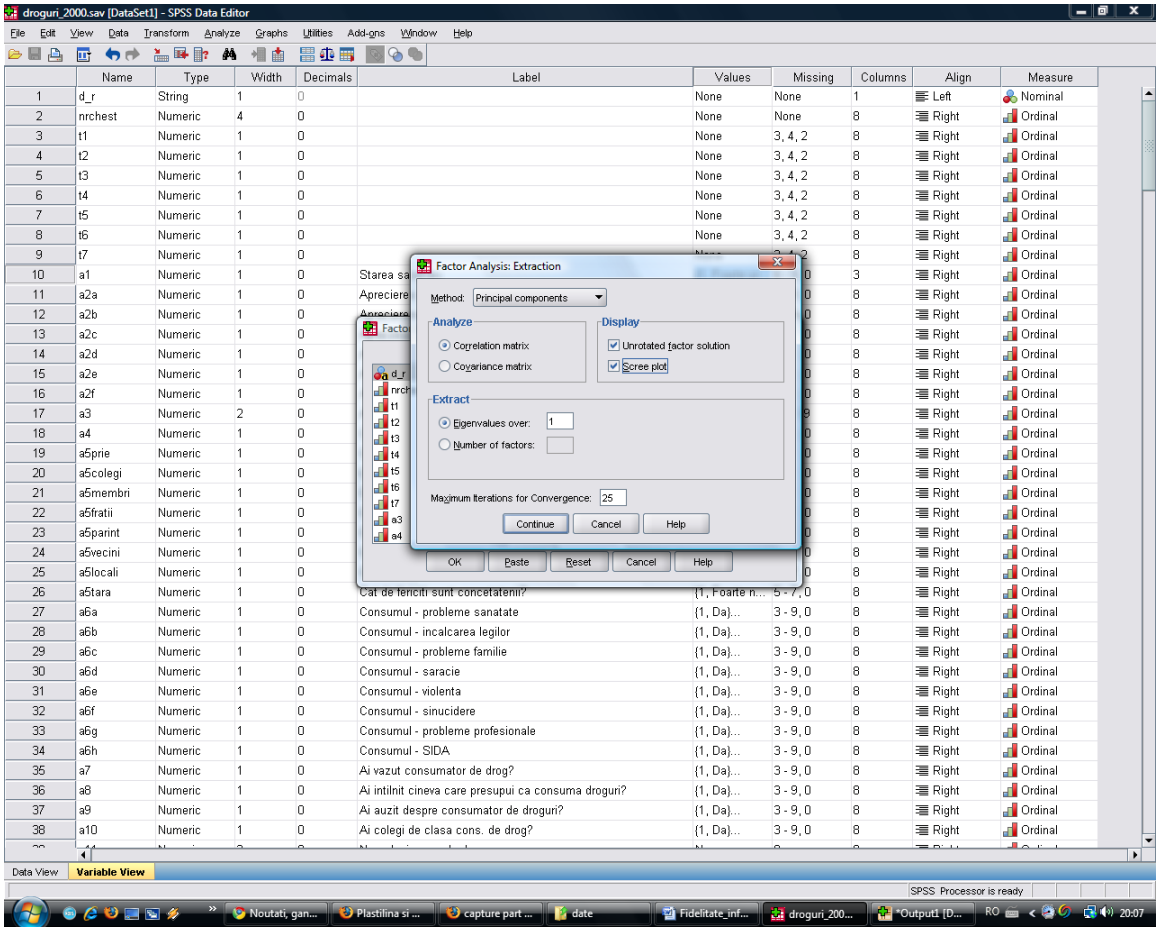
Selectați variabilele a1-a2f din lista de variabile și le transferați în fereastra *Variables.*



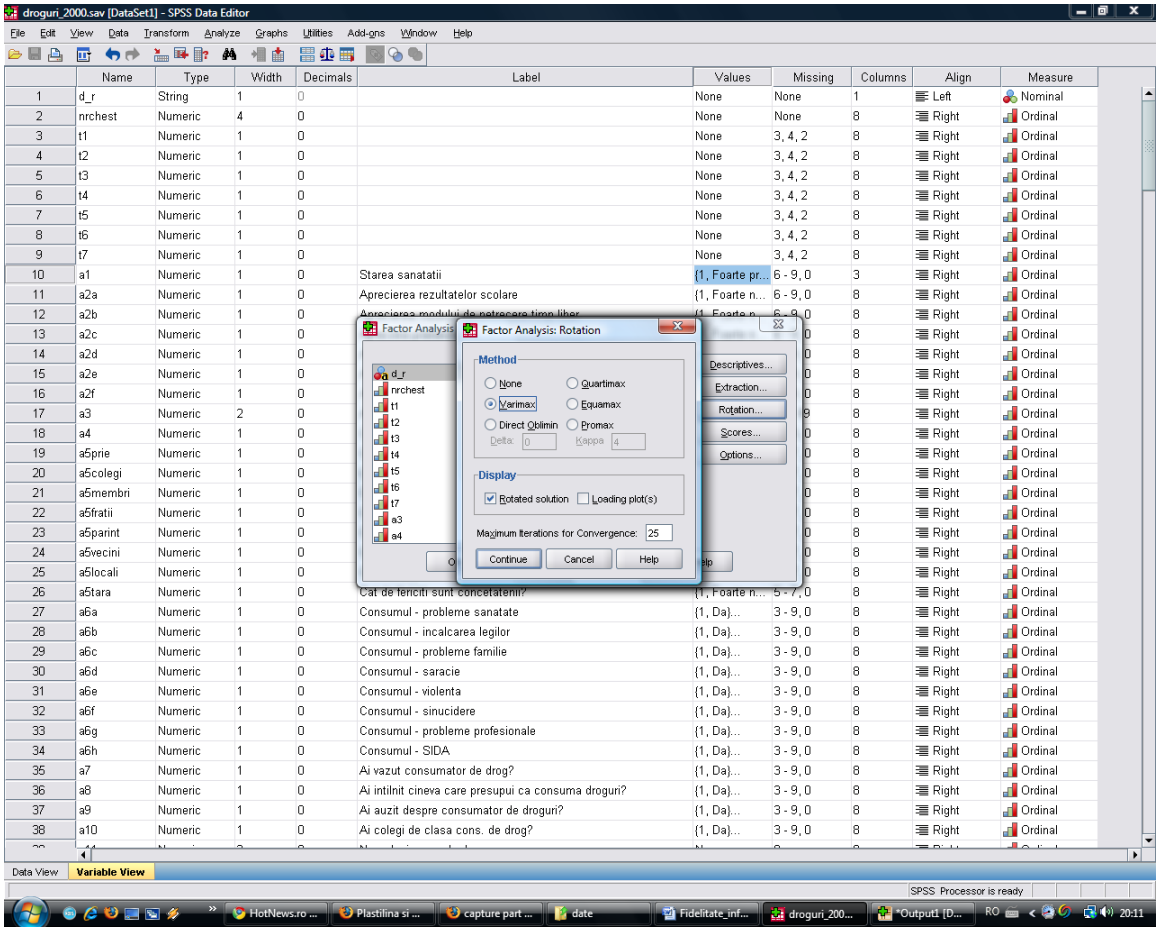
Apăsați butonul Extraction.
Din fereastra de dialog care se deschide selectați prima opțiune la Method *(principal components)*.
Să fie marcate următoarele opțiuni:

- *Correlation matrix*
- *Unrotated factor solution*
- *Scree plot*
- *Eigenvalues over 1*

Apăsați **Continue**, apoi din fereastra de dialog principală apăsați butonul **Rotation**.



În fereastra de dialog care se deschide selectați *Varimax* la *Method* și *Rotated solution* la *Display*. Apăsați Continue.



Din rezultatul scris în output, un tabel important este cel cu Total Variance Explained. Rezultatul ne arată că cu excepția unui factor, nici unul nu explică mai mult decât varianța unei variabile (toate *eigenvalues* sunt mai mici de 1 în afară de prima componentă).

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	2,321	33,156	33,156	2,321	33,156	33,156
2	,972	13,889	47,044			
3	,906	12,950	59,994			
4	,796	11,371	71,366			
5	,726	10,368	81,734			
6	,673	9,617	91,351			
7	,605	8,649	100,000			

Extraction Method: Principal Component Analysis.

Rezultatul sugerează că scala compusă din cei 7 itemi este unidimensională. Factorul 1, cel extras, explică 33,15% din variația celor 7 itemi.

Un alt rezultat important în acest sens este tabelul saturațiilor primei componente (care reflectă dimensiunea măsurată probabil de scală) în itemi. Acești coeficienți, care au o variabilitate similară cu cea a celor de corelație sunt destul de uniformi

Component Matrix[a]

	Component 1
Starea sanatatii	,554
Aprecierea rezultatelor scolare	,415
Aprecierea modului de petrecere timp liber	,646
Aprecirea prietenilor	,583
Aprecierea vietii sentimentale	,577
Aprecierea posibilitatilor materiale	,638
Aprecierea relatiilor cu parintii	,586

Extraction Method: Principal Component Analysis.

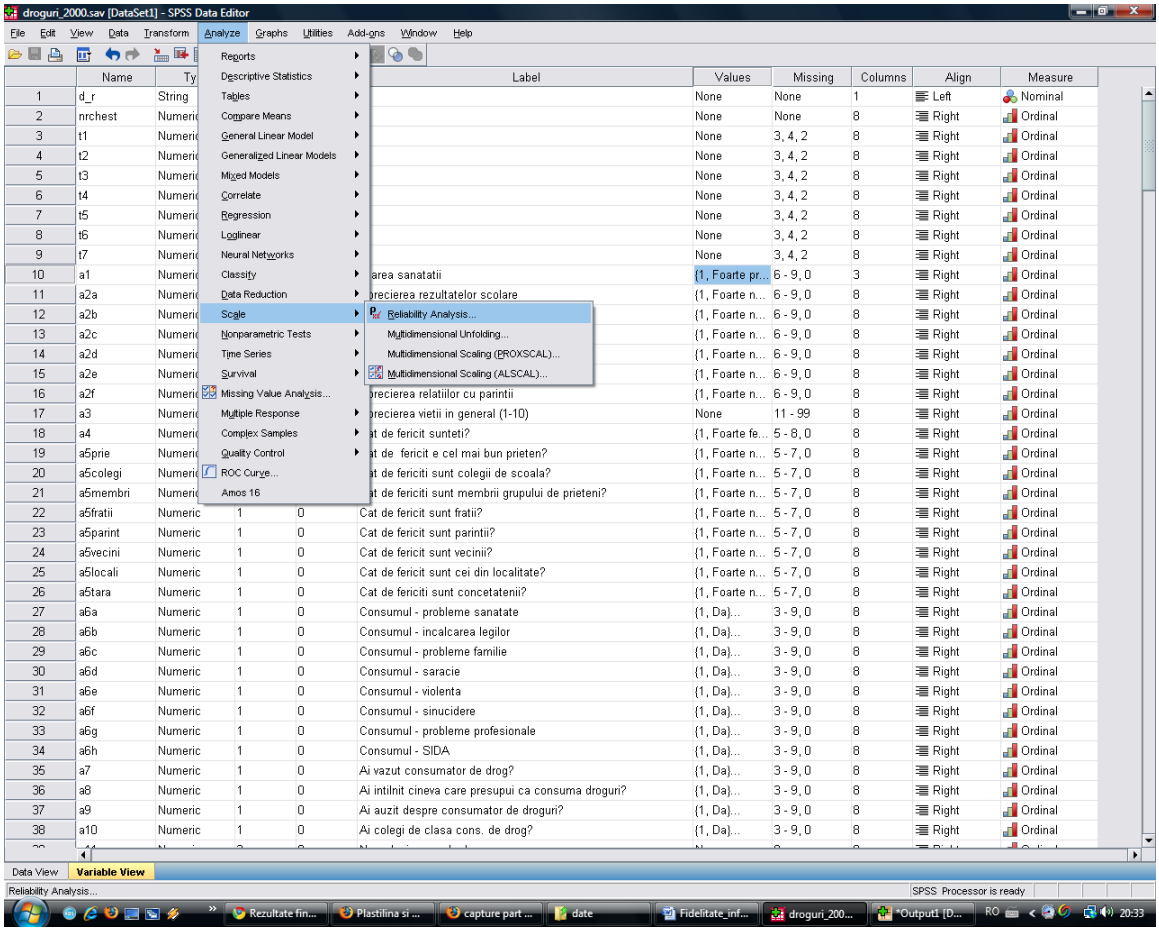
a. 1 components extracted.

Concluzie: scala este unidimensională, analiza factorială extrăgând doar un factor. Tabelul saturațiilor sugerează posibile probleme de consistență ale itemului *Aprecierea rezultatelor școlare* (care are saturația cea mai mică).

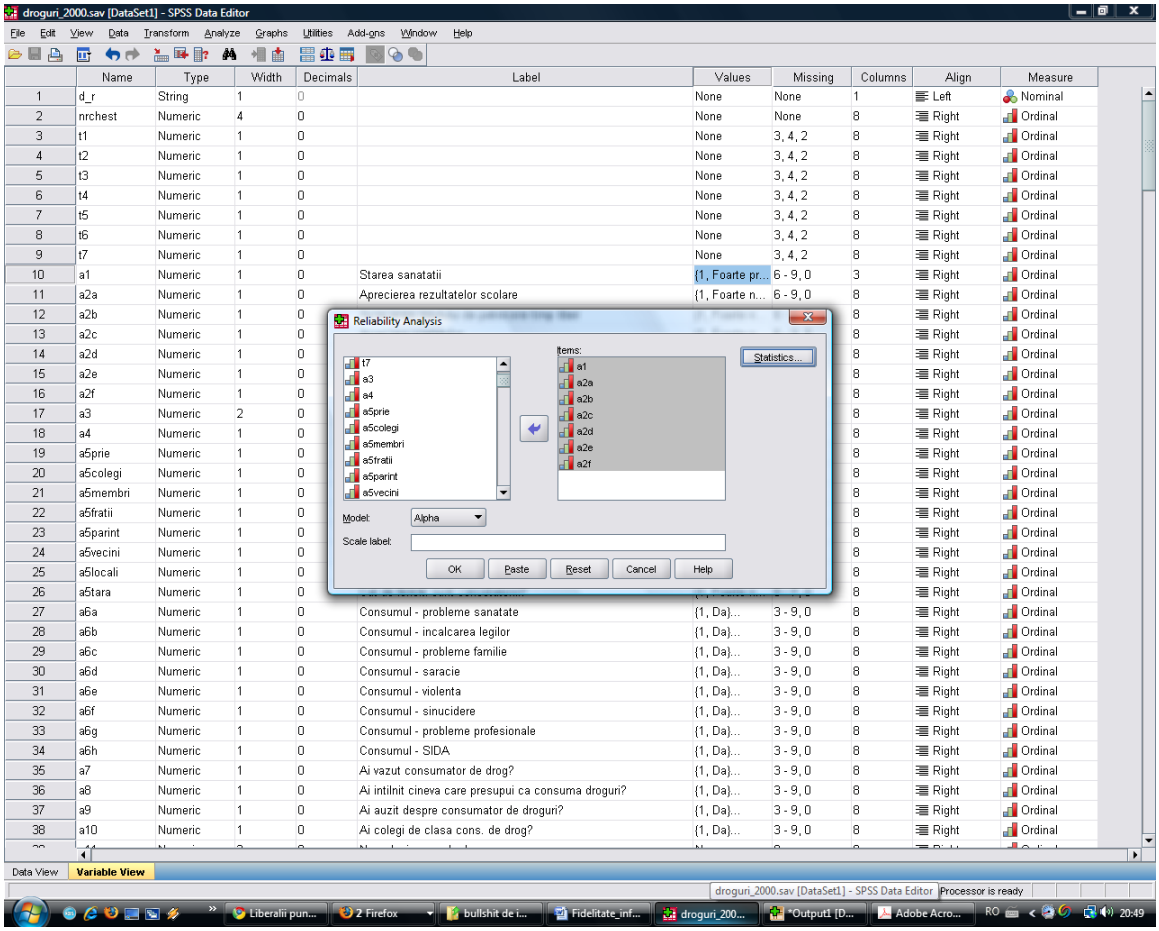
Analiza fidelității

Model alpha Cronbach

Selectăm opțiunea Reliability Analysis din **Analyze →Scale**



În fereastra de dialog selectăm itemii în stânga și în mutăm la **Items**.
La meniul derulant **Model** ne asigurăm că este selectată opțiunea **Alpha**.
Apăsăm butonul **Statistics.**

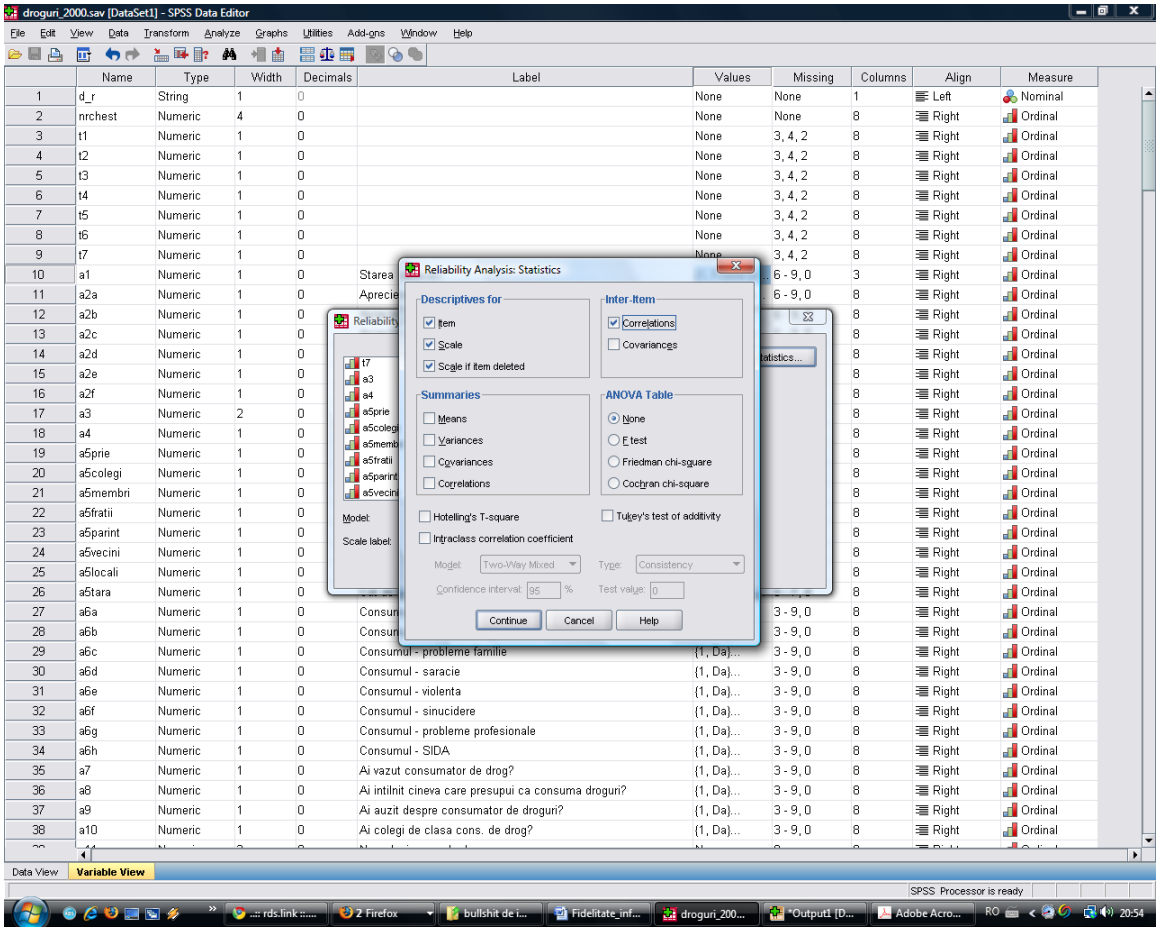


	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
1	d_r	String	1	0		None	None	1	Left	Nominal
2	nrchest	Numeric	4	0		None	None	8	Right	Ordinal
3	t1	Numeric	1	0		None	3, 4, 2	8	Right	Ordinal
4	t2	Numeric	1	0		None	3, 4, 2	8	Right	Ordinal
5	t3	Numeric	1	0		None	3, 4, 2	8	Right	Ordinal
6	t4	Numeric	1	0		None	3, 4, 2	8	Right	Ordinal
7	t5	Numeric	1	0		None	3, 4, 2	8	Right	Ordinal
8	t6	Numeric	1	0		None	3, 4, 2	8	Right	Ordinal
9	t7	Numeric	1	0		None	3, 4, 2	8	Right	Ordinal
10	a1	Numeric	1	0	Starea sanatatii	{1, Foarte pr...	6 - 9, 0	3	Right	Ordinal
11	a2a	Numeric	1	0	Aprecierea rezultatelor scolare	{1, Foarte n...	6 - 9, 0	8	Right	Ordinal
12	a2b	Numeric	1	0				8	Right	Ordinal
13	a2c	Numeric	1	0				8	Right	Ordinal
14	a2d	Numeric	1	0				8	Right	Ordinal
15	a2e	Numeric	1	0				8	Right	Ordinal
16	a2f	Numeric	1	0				8	Right	Ordinal
17	a3	Numeric	2	0				8	Right	Ordinal
18	a4	Numeric	1	0				8	Right	Ordinal
19	a5prie	Numeric	1	0				8	Right	Ordinal
20	a5colegi	Numeric	1	0				8	Right	Ordinal
21	a5membri	Numeric	1	0				8	Right	Ordinal
22	a5fratii	Numeric	1	0				8	Right	Ordinal
23	a5parint	Numeric	1	0				8	Right	Ordinal
24	a5vecini	Numeric	1	0				8	Right	Ordinal
25	a5locali	Numeric	1	0				8	Right	Ordinal
26	a5tara	Numeric	1	0				8	Right	Ordinal
27	a6a	Numeric	1	0	Consumul - probleme sanatate	{1, Da}...	3 - 9, 0	8	Right	Ordinal
28	a6b	Numeric	1	0	Consumul - incalcarea legilor	{1, Da}...	3 - 9, 0	8	Right	Ordinal
29	a6c	Numeric	1	0	Consumul - probleme familie	{1, Da}...	3 - 9, 0	8	Right	Ordinal
30	a6d	Numeric	1	0	Consumul - saracie	{1, Da}...	3 - 9, 0	8	Right	Ordinal
31	a6e	Numeric	1	0	Consumul - violenta	{1, Da}...	3 - 9, 0	8	Right	Ordinal
32	a6f	Numeric	1	0	Consumul - sinucidere	{1, Da}...	3 - 9, 0	8	Right	Ordinal
33	a6g	Numeric	1	0	Consumul - probleme profesionale	{1, Da}...	3 - 9, 0	8	Right	Ordinal
34	a6h	Numeric	1	0	Consumul - SIDA	{1, Da}...	3 - 9, 0	8	Right	Ordinal
35	a7	Numeric	1	0	Ai vazut consumator de drog?	{1, Da}...	3 - 9, 0	8	Right	Ordinal
36	a8	Numeric	1	0	Ai intilnit cineva care presupui ca consuma droguri?	{1, Da}...	3 - 9, 0	8	Right	Ordinal
37	a9	Numeric	1	0	Ai auzit despre consumator de droguri?	{1, Da}...	3 - 9, 0	8	Right	Ordinal
38	a10	Numeric	1	0	Ai colegi de clasa cons. de drog?	{1, Da}...	3 - 9, 0	8	Right	Ordinal

Solicităm în mediul Statistics:
Descriptives for:

- Item
- Scale
- Scale if item deleted

Inter-item Correlations
Apăsăm **Continue**



Rezultatele analizei de fidelitate
Analiza a inclus 938 de cazuri neponderate și a exclus 62 de cazuri care aveau cel puțin un caz lipsă (excludere listwise).

[DataSet1] D:\lucru\paginawebscurta\date\droguri_2000.sav

Scale: ALL VARIABLES

Case Processing Summary

		N	%
Cases	Valid	938	93,8
	Excluded[a]	62	6,2
	Total	1000	100,0

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
,657	,657	7

Item Statistics

	Mean	Std. Deviation	N
Starea sanatatii	3,85	,822	938
Aprecierea rezultatelor scolare	3,35	,878	938
Aprecierea modului de petrecere timp liber	3,75	1,031	938
Aprecirea prietenilor	4,00	,969	938
Aprecierea vietii sentimentale	3,39	1,215	938
Aprecierea posibilitatilor materiale	3,52	1,019	938
Aprecierea relatiilor cu parintii	4,01	1,066	938

Inter-Item Correlation Matrix

	Starea sanatatii	Aprecierea rezultatelor scolare	Aprecierea modului de petrecere timp liber	Aprecirea prietenilor	Aprecierea vietii sentimentale	Aprecierea posibilitatilor materiale	Aprecierea relatiilor cu parintii
Starea sanatatii	1,000	,120	,180	,227	,160	,260	,269
Aprecierea rezultatelor scolare	,120	1,000	,161	,133	,167	,143	,188
Aprecierea modului de petrecere timp liber	,180	,161	1,000	,321	,344	,259	,225
Aprecirea prietenilor	,227	,133	,321	1,000	,220	,229	,182
Aprecierea vietii sentimentale	,160	,167	,344	,220	1,000	,255	,156
Aprecierea posibilitatilor materiale	,260	,143	,259	,229	,255	1,000	,317
Aprecierea relatiilor cu parintii	,269	,188	,225	,182	,156	,317	1,000

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
Starea sanatatii	22,01	13,511	,342	,134	,630
Aprecierea rezultatelor scolare	22,52	13,827	,253	,067	,651
Aprecierea modului de petrecere timp liber	22,11	12,053	,438	,214	,599

Aprecirea prietenilor	21,87	12,737	,372	,156	,620
Aprecierea vietii sentimentale	22,47	11,686	,372	,166	,623
Aprecierea posibilitatilor materiale	22,35	12,208	,422	,191	,604
Aprecierea relatiilor cu parintii	21,85	12,332	,372	,168	,620

Scale Statistics			
Mean	Variance	Std. Deviation	N of Items
25,86	16,252	4,031	7

Interpretarea scorului alpha Cronbach

Scorul alpha (α) se bazează pe corelația medie dintre itemi. Formula lui α este:

$\alpha=kr/(1+(k-1)r)$

Unde k reprezintă numărul de itemi iar r este media coeficienților de corelație dintre itemi. Din formulă se înțelege că mărimea lui α depinde liniar și pozitiv atât de numărul de itemi cât și de corelațiile dintre itemi. Scorul nostru de 0,659 este unul relativ satisfăcător. Convenția statistică cere un α de minimum 0,7 pentru o fidelitate certă.

Tabelul de mai jos permite evaluarea efectului fiecărui item asupra scorului total al scalei. Prima coloană cuprinde media scalei în cazul în care itemul de pe linia respectivă ar fi exclus din calculul scorului.

Coloana a doua cuprinde varianța totală a scalei în cazul eliminării itemului de pe linia respectivă.

Coloana a treia (Corrected item-total correlation) arată corelația dintre item și scorul calculat fără itemul respectiv: observăm că itemul al doilea, referitor la aprecierea rezultatelor școlare, are corelația cea mai mică cu restul scalei.

O altă modalitate de apreciere a consistenței relațiilor itemilor cu scala este regresia multiplă a fiecărui item pe baza celorlalți itemi, considerați variabile independente. Coloana a patra (squared multiple correlation) cuprinde coeficienții R^2 de determinare ai acestor regresii: remarcăm din nou că relația cea mai slabă cu scala este a celei de-a doua variabile, în timp ce „aprecierea modului de petrecere a timpului liber” are relația cea mai consistentă.

Ultima coloană ne arată valoarea lui α în cazul în care itemul este exclus din construcția scalei.

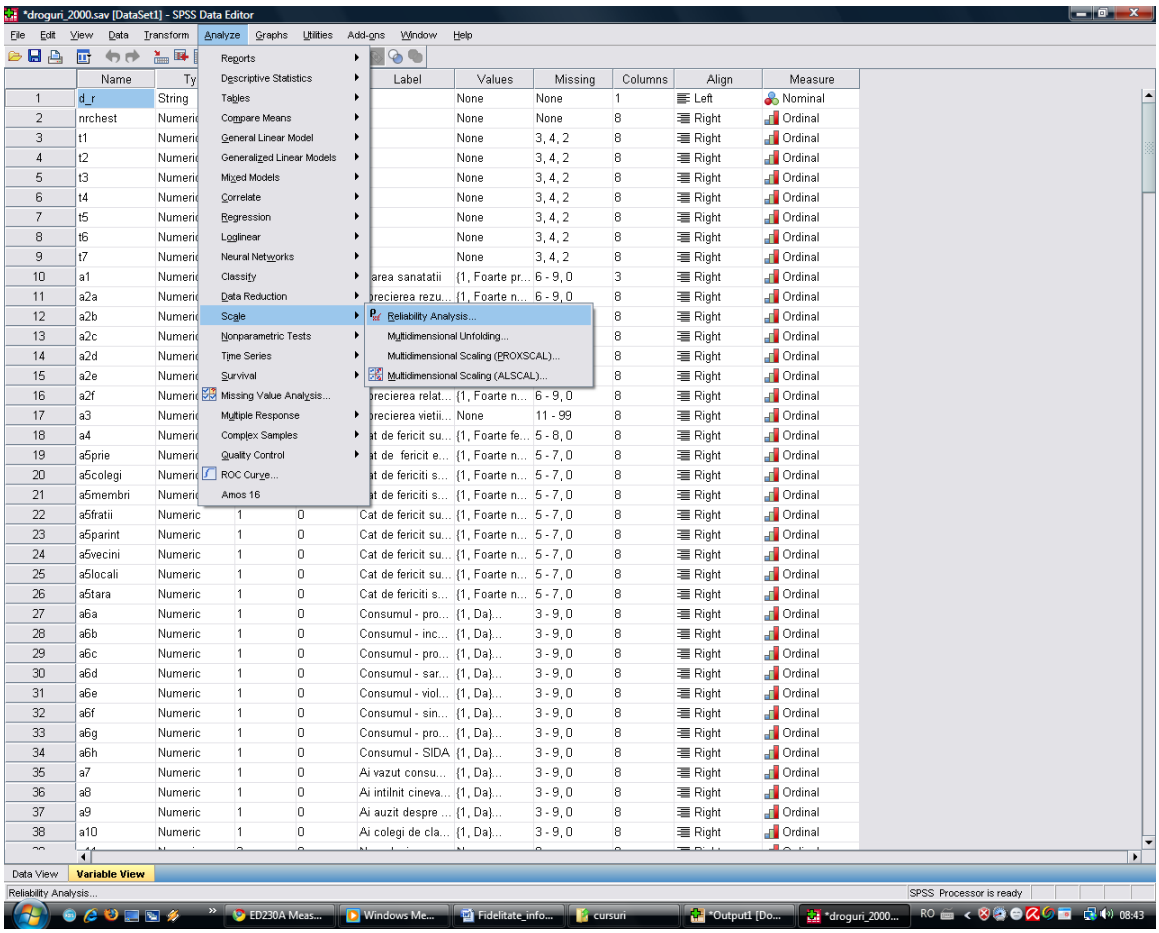
Valorile din tabel ne indică faptul că eliminarea oricărui item duce la scăderea coeficientului de fidelitate.

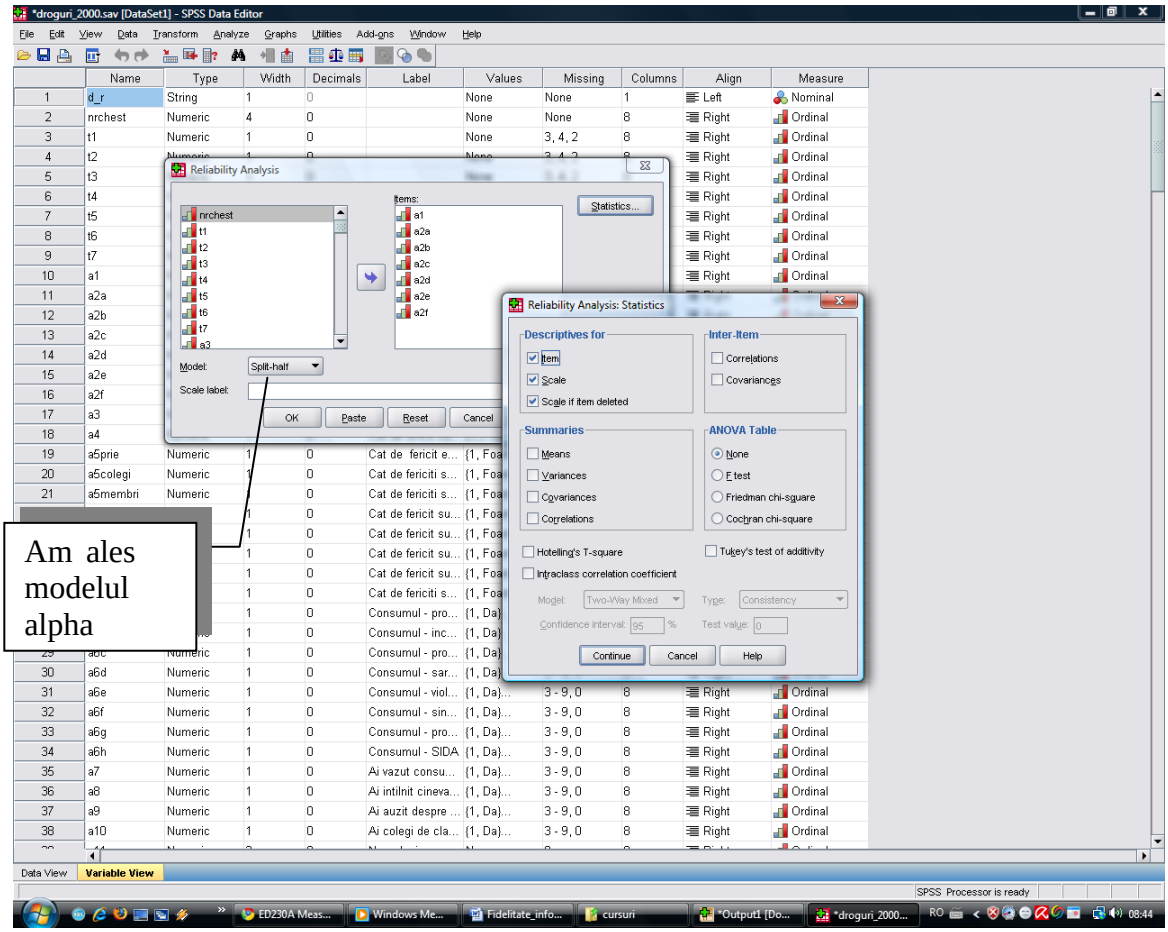
Prin urmare, nu se justifică scăderea numărului de itemi.

Modelul calculului fidelității prin înjumătățire (split half)

- Corelaţia dintre cele două forme (rj)
- Coeficientul Speaman- Brown - măsură care profeţeşte fidelitatea întregului test, pe baza corelaţiilor dintre două jumătăţi de test (2*rj/(1+rj)). Cele două jumătăţi de test trebuie să aibă varianţe egale. Pentru asta, e bine ca itemii să fie selectaţi aleator in scale. De fapt asta este una dintre neajunsurile metodei: faptul că rezultatele depind de modul de împărțire a itemilor în jumătățile de scale.
- Coeficientul Guttman de fidelitate split-half - ce şi cel de mai sus, doar că nu se presupune că varianţele celor două jumătăţi de scală ar fi egale.

Pașii care trebuie urmați în SPSS sunt identici cu cei de la modelul alpha, doar că se schimbă modelul.





După selectarea opțiunilor în fereastra Realibility Analysis: Statistics, apăsăm Continue apoi OK. Rezultatul exportat în html este mai jos:

Scale: ALL VARIABLES

Case Processing Summary

		N	%
Cases	Valid	938	93,8
	Excluded[a]	62	6,2
	Total	1000	100,0

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	Part 1	Value	,489
		N of Items	4,000[a]
	Part 2	Value	,481
		N of Items	3,000[b]
	Total N of Items		7,000

Întrucât alpha depinde de numărul de itemi, era de așteptat o scădere a coeficientului pentru jumătățile de itemi.

Correlation Between Forms		,502
Spearman-Brown Coefficient	Equal Length	,668
	Unequal Length	,672
Guttman Split-Half Coefficient		,668
a. The items are: Starea sanatatii , Aprecierea rezultatelor scolare, Aprecierea modului de petrecere timp liber, Aprecirea prietenilor.		
b. The items are: Aprecirea prietenilor, Aprecierea vietii sentimentale, Aprecierea posibilitatilor materiale, Aprecierea relatiilor cu parintii.		

Întrucât scalele au lungimi inegale citim acest coeficient. El este nesatisfăcător, fiind mai mic de 0,7

Coeficientul lui Guttman este și el mai mic de 0,7

Item Statistics

	Mean	Std. Deviation	N
Starea sanatatii	3,85	,822	938
Aprecierea rezultatelor scolare	3,35	,878	938
Aprecierea modului de petrecere timp liber	3,75	1,031	938
Aprecirea prietenilor	4,00	,969	938
Aprecierea vietii sentimentale	3,39	1,215	938
Aprecierea posibilitatilor materiale	3,52	1,019	938
Aprecierea relatiilor cu parintii	4,01	1,066	938

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
Starea sanatatii	22,01	13,511	,342	,630
Aprecierea rezultatelor scolare	22,52	13,827	,253	,651
Aprecierea modului de petrecere timp liber	22,11	12,053	,438	,599
Aprecirea prietenilor	21,87	12,737	,372	,620
Aprecierea vietii sentimentale	22,47	11,686	,372	,623
Aprecierea posibilitatilor materiale	22,35	12,208	,422	,604
Aprecierea relatiilor cu parintii	21,85	12,332	,372	,620

Scale Statistics

	Mean	Variance	Std. Deviation	N of Items
Part 1	14,95	5,446	2,334	4[a]
Part 2	10,92	5,373	2,318	3[b]
Both Parts	25,86	16,252	4,031	7
a. The items are: Starea sanatatii , Aprecierea rezultatelor scolare, Aprecierea modului de petrecere timp liber, Aprecirea prietenilor.				
b. The items are: Aprecierea vietii sentimentale, Aprecierea posibilitatilor materiale, Aprecierea relatiilor cu parintii.				

SPSS împarte scala în două jumătăți. Întrucât avem un număr impar de itemi, primii patru i-a alocat primei scale iar restul de trei au compus a doua scală. REstul rezultatelor sunt idetice cu modelul alpha.