

# SPSS

**Statistical Package for the Social Sciences**

Introducerea datelor

Editorul de date este asemănător cu cel folosit de programele de calcul tabelar (Excel ș.a.) pentru foile de lucru. Avem aici două vederi diferite:

- Data View: se folosește pentru introducerea datelor
- Variable View: este utilizată pentru definirea variabilelor

Fiecare rând reprezintă o variabilă iar fiecare coloană câte un atribut al fiecărei variabile.

Avem următoarele attribute care trebuie specificate pentru fiecare variabilă:

1. Numele (**NAME**) – maximum 8 caractere pentru variantele mai vechi de program, 64 de caractere pentru cele mai noi; numele trebuie să fie unice și nu pot include spații.
2. Tipul (**TYPE**) – cel mai adesea se folosim tipul numeric pentru că la variabilele măsurate la nivel nominal sau ordinal pre-definim valorile pe care acestea le pot lua și vom introduce doar numărul corespunzător valorii respective iar la nivel interval sau rapoarte vorbim despre cifre. Tipul **STRING**(șir de caractere) se folosește pentru răspunsurile libere care conțin și altceva decât cifre.

3. Dimensiunea (**WIDTH**) – ne spune câte caractere putem introduce în variabila noastră. În mod implicit pentru **Numeric** avem 8, din care 2 după virgulă (**Decimals**)
4. Eticheta (**LABEL**) – ne permite să precizăm pe larg despre ce este vorba în variabila respectivă
5. Valorile (**VALUES**)- se folosesc pentru realizarea unor codificări (variantele de răspuns din chestionar) care ne ajută să simplificăm procedura de introducere. *Exemplu:* Pentru variabila genul persoanei putem codifica cu 1 valoarea „feminin”, cu 2 valoarea „masculin”. Când vom introduce datele vom introduce numărul variantei alese.

6. Valori lipsă (**MISSING**)= este o opțiune care se folosește atunci când dorim să fie eliminate din calcule anumite valori (cum ar fi *Nu știu*, *Nu răspund*, *Nu este cazul* ș.a.). Pentru a lăsa anumite valori în afara calculului, alegem **discrete missing values**.  
*Exemplu:* Non-răspunsurile sunt codificate cu 99 sau 77
7. Dimensiunea coloanei (**COLUMNS**)– se folosește pentru a specifica numărul de caractere folosit pentru a afișa valorile variabilei respective.
8. Aliniamentul (**ALIGN**) – tot pentru afișarea datelor trebuie să specificăm modul de aliniere al valorilor pentru fiecare variabilă – la stânga, la dreapta sau centrat.

9. **Nivelul de măsurare (MEASURE)**, care are trei posibile valori: nominală, ordinală și scale (care include interval și rapoarte).
10. În același scop, cel de a ușura a ușura selecția variabilelor pentru diferite proceduri statistice putem specifica și rolul pe care-l va îndeplini fiecare variabilă: Input (variabilă independentă), Target (variabilă dependentă), Both (ambele) sau None (unde nu avem un rol anume).

După finalizarea selectării variabilelor și a coeficienților și indicatorilor pentru orice procedură statistică, după ce dăm **OK** și se deschide o nouă fereastră – **OUTPUT** unde găsim rezultatul analizei SPSS.

Frecvențele



- Pasii care trebuie efectuati: Din meniul programului selectam:
- **ANALYZE** → **DESCRIPTIVE STATISTICS** → **FREQUENCIES** → butonul **STATISTICS** adaugam variabilele cu ajutorul sagetii de pe ecran sau prin dublu click.
- Apasam butonul **Statistics** si selectam indicatorii doriti (in functie de nivelul de masurare al variabilelor putem alege intre Mean, Median, Mode – pentru tendinta centrala; si Std. Deviation, abaterea standard pentru imprastiere.
- → Butonul Charts poate fi utilizat pentru generarea unor **grafice** de tip bar – cu bare, **pie** – placinta (sau sector) – pentru variabile masurate la nivel nominal sau ordinal si **histograms** – histograme, pentru variabilele de tip scale.

- **Indicatorii tendinței centrale** (sau de poziție) cei mai des utilizați sunt:

- a) **modul**

- b) **mediana**

- c) **media.**

## **Amplitudinea**

*Ex: putem avea note între 1-10*

Amplitudinea:  $10 - 1 = 9 \Rightarrow$  diferența dintre valoarea cea mai mare și cea mai mică

- **Modul** este definit ca fiind valoarea cu frecvența cea mai mare a unei distribuții. Altfel spus, modul este acea valoare a variabilei care apare cel mai des într-un eșantion sau într-o populație.
- **Mediana** este definită ca fiind acea valoare a unei variabile care împarte seria ordonată de date în două părți egale, astfel încât 50% din observații se vor situa deasupra valorii mediane iar 50% dedesubtul ei.
- **Media** (aritmetică) este probabil cea mai importantă și totodată cea mai populară măsură a tendinței centrale a unei distribuții. Ea se calculează ca sumă a tuturor valorilor observate ale distribuției împărțită la numărul total de observații

# Testarea ipotezelor

# Asocierea

- Se foloseste pentru variabilele nominale si/sau ordinale
- PASI in SPSS:
  - **ANALYZE → DESCRIPTIVE STATISTICS → CROSSTABS**
  - Adaugam variabilele si pe randuri si pe coloane folosind sagetile corespunzatoare
  - Apasam butonul **Statistics** si selectam coeficientii doriti (Chi-Square si, in functie de nivelul de masurare al variabilelor, Lambda pentru nominal respectiv Gamma pentru ordinal)

# Interpretare

- 1. Din tabelul de asociere (**crosstabulation**) interpretăm cele mai relevante valori (pe rând și pe coloană)

Cât de mulțumit(a) sunteți în general de felul în care trăiți? \* Sexul respondentului

## Crosstabulation

Count

		Sexul respondentului		Total
		masculin	feminin	
Cât de mulțumit(ă) sunteți în general de felul în care trăiți?	Deloc mulțumit	133	178	311
	Nu prea mulțumit	437	527	964
	Destul de mulțumit	333	328	661
	Foarte mulțumit	20	28	48
	NS	3	8	11
	NR	1	4	5
Total		927	1073	2000

- **2. Semnificația lui  $hi^2$** , care apare pe ultima coloană a tabelului **chi-square tests**, de obicei în dreptul lui *pearson chi-square*. Pentru a avea o legătură semnificativă între cele două variabile, **semnificația lui  $hi^2$  trebuie să fie sub 0,05**.

Tabel 15: Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	16.100 <sup>a</sup>	7	.024
Likelihood Ratio	21.952	7	.003
Linear-by-Linear Association	10.330	1	.001
N of Valid Cases	21		

a. 16 cells (100.0%) have expected count less than 5. The minimum expected count is .43.

### 3. Interpretăm **gamma** sau **lambda** din tabelul **symmetric measures**:

- interpretăm valoarea absolută a lui **gamma** (**value**)
- semnul valorii lui **gamma**
- semnificația lui **gamma**

Tabel 16: Symmetric Measures

		Value	Asymp. Std. Error <sup>a</sup>	Approx. T <sup>b</sup>	Approx. Sig.
Ordinal by Ordinal	Gamma	-.922	.072	-7.974	.000
N of Valid Cases		21			

a. Not assuming the null hypothesis.

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



- → **gamma** poate lua valori între -1 și 1. Dacă valoarea ei este 1 sau -1, atunci există o relație foarte puternică între cele două variabile. Cu cât valoarea lui **gamma** este mai apropiată de 0, fie că e negativă, fie că e pozitivă, cu atât legătura dintre cele două variabile va fi mai slabă. Valoarea absolută exprimă tăria legăturii între cele două variabile.

Tabel 16: Symmetric Measures

		Value	Asymp. Std. Error <sup>a</sup>	Approx. T <sup>b</sup>	Approx. Sig.
Ordinal by Ordinal	Gamma	-.922	.072	-7.974	.000
N of Valid Cases		21			

a. Not assuming the null hypothesis.

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

- semnul lui **gamma** (+; -) exprimă direcția relației dintre cele două variabile. Dacă semnul e „+”, atunci există o relație pozitivă între variabile: **relația pozitivă** = cele două variabile vor varia – vor crește sau descrește împreună. Dacă **gamma** este negativ, **relația** va fi **negativă** = una va crește, pe când cealaltă va scădea și invers.

Tabel 16: Symmetric Measures

		Value	Asymp. Std. Error <sup>a</sup>	Approx. T <sup>b</sup>	Approx. Sig.
Ordinal by Ordinal	Gamma	-.922	.072	-7.974	.000
N of Valid Cases		21			

a. Not assuming the null hypothesis.

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

- semnificația lui **gamma** (ultima coloană – **approx. sig.**) confirmă sau nu semnificația lui **hi<sup>2</sup>**. Semnificația trebuie să fie **sub 0,05** pentru a confirma relația certificată de semnificația lui **hi<sup>2</sup>**. **Semnificația lui hi<sup>2</sup> este cea mai importantă** – dacă din ea rezultă ca nu există o legătură între variabile, **gamma** sau **lambda** nu se mai interpretează.

Tabel 16: Symmetric Measures

		Value	Asymp. Std. Error <sup>a</sup>	Approx. T <sup>b</sup>	Approx. Sig.
Ordinal by Ordinal	Gamma	-.922	.072	-7.974	.000
N of Valid Cases		21			

a. Not assuming the null hypothesis.

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

- → **lambda** nu are semn. Poate lua valori intre 0 si 1, cu cat valoarea este mai mare cu atat relatia este mai puternica. Semnificatia se interpreteaza asemanator cu cea a lui gamma

# Regresia

- Se foloseste pentru variabilele de tip scala
- Modelul de regresie este format dintr-o variabila dependenta si doua sau mai multe independente
- PASI in SPSS:
  - **ANALYZE → REGRESSION → LINEAR**
  - Selectam variabila dependenta si variabilele independente cu ajutorul sagetilor:

# Interpretare

- 1. Îl interpretăm pe **R<sup>2</sup> (R Square)** din tabelul **model summary**, care ne arată cât la sută din variația variabilelor dependente este explicată de variația variabilelor independente. Dacă **R<sup>2</sup>** este 0,246, atunci 24,6% din variația variabilei dependente este explicată de către variația variabilelor independente. **R<sup>2</sup>** poate lua valori între 0 și 1. Dacă valoarea lui **R<sup>2</sup>** este 1, atunci 100% din variația variabilelor dependente este explicată de variația variabilelor independente.

Tabel 2: Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.695 <sup>a</sup>	<b>.483</b>	.450	2.106

a. Predictors: (Constant), Venitul familiei dumneavoastra pe ultima luna este, In acest moment venitul dumneavoastra net este

- 2. Interpretăm **semnificația modelului de regresie – ANOVA** – se regăsește în ultima coloană, fiind primul rând. Dacă este sub 0,05, atunci modelul de regresie este semnificativ.

Tabel 3: ANOVA<sup>b</sup>

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	128.582	2	64.291	14.500	<b>.000<sup>a</sup></b>
	Residual	137.447	31	4.434		
	Total	266.029	33			

a. Predictors: (Constant), Venitul familiei dumneavoastra pe ultima luna este, In acest moment venitul dumneavoastra net este

b. Dependent Variable: Pe o scala de la 1-10 in ce masura considerati ca absolvirea masterului v-a schimbat nivelul de trai?

- **3. Coeficienții standardizați Beta** (a II-a coloană) ne arată puterea explicativă a fiecărei variabile independente. Cu cât valoarea absolută este mai mare, neținându-se cont de semn, cu atât puterea explicativă a variabilei respective este mai mare. Semnul lui Beta ne spune ce fel de relație există între variabila respectivă și cea dependentă, pozitivă sau negativă. Semnificația (Sig.) a coeficienților Beta trebuie să respecte regula cunoscută (să fie  $<.05$ ) pentru a putea spune că respectiva variabilă independentă influențează variabila dependentă.

Tabel 10: Coefficients<sup>a</sup>

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.		
	B	Std. Error	Beta				
1	(Constant)	-14.984	7.130		-2.102	.044	
	Media de absolvire a programului de masterat	1.235	.560		.362	2.204	.036
	Varsta	.364	.218		.274	1.666	.107

a. Dependent Variable: Pe o scala de la 1-10 in ce masura considerati ca absolvirea masterului v-a schimbat nivelul de trai?