



Analiza datelor de marketing utilizand S.P.S.S.

- curs introductiv -

Mihai Orzan

- mihai.orzan@ase.ro
- joi, 19:30, sala 1406



Chestiuni organizatorice

- **Nota:**
 - Examen final (1 iunie): 40%
 - Test seminar: 60%
- <http://orzanm.ase.ro/spss>
- **Suport curs:**
 - Cătoiu I. (coord.), Bălan C., Dănețiu T., Orzan Gh., Popescu I., Vegheș C., Vrânceanu D. - "*Cercetări de marketing*", Ed. Uranus, 2002
 - Popa, D., "*Analiza datelor in psihologie. Teorie si practica cu SPSS*", Ed. Polirom, 2008.
- **Suport semiar (recomandat):**
 - Howitt, D. si Cramer, D., "*Introducere in SPSS pentru psihologie*", Ed. Polirom, 2006.



Ce reprezinta analiza datelor?



Analiza datelor reprezinta un proces complex si sistematic de aplicare a tehnicilor statistico-matematice, in scopul extragerii din colectia de date constituita a informatiilor necesare procesului decizional



Surse de date



- **Surse de date:**
 - Cercetari cantitative:
 - Sondaje;
 - Observari;
 - Surse secundare de date:
 - interne;
 - externe;
 - Cercetari calitative:
 - focus grupuri;
 - clientul misterios;
 - interviuri in profunzime;
 - teste de utilitate;





Clasificare tipurilor de analiza

- Numarul esantioanelor cercetate
 - ➡ Un singur esantion;
 - ➡ Doua esantioane;
 - ➡ Mai mult de doua esantioane;
- Natura relatiei dintre esantioane
 - ➡ Esantioane independente;
 - ➡ Esantioane dependente;





Clasificare tipurilor de analiza

- Numarul variabilelor utilizate:
 - ➡ O singura variabila (**analiza univariata**);
 - ➡ Doua variabile (**analiza bivariata**);
 - ➡ Mai mult de doua variabile (**analiza multivariata**);
- Natura relatiei dintre variabile
 - ➡ Variabile independente;
 - ➡ Variabile dependente;





Obiectivele analizei

- Determinarea **tendinței centrale**;
- Caracterizarea variației și dispersiei;
- Măsurarea gradului de **asociere**;
- Realizarea de estimări și **previziuni**;
- Evaluarea semnificației **diferențelor** dintre variabile și grupuri de variabile;
- Evidențierea legăturilor **cauzale**;





Obiectivele analizei

- Analiza primara a datelor
 - Masurarea tendintei centrale
 - Analiza variatiei
 - Stabilirea normalitatii distributiei
- Analiza diferentiala
 - Identificarea existentei unor diferente statistice intre esantioane (grupuri) de respondenti
- Analiza asociativa
 - Identificarea existentei unor asemanari (asocieri) semnificative statistic variabile demografice si psiho-comportamentale





Obiectivele analizei

Analiza inferentiala

- Identificarea gradului in care valorile identificate la nivel de esantion sunt reprezentative la nivelul populatiilor investigate
- Analiza predictiva
 - Identificarea evolutiei ulterioare a fenomenelor investigate
- Analiza complexa a datelor
 - Analiza canonica, analiza discriminantului, analiza structurilor latente
 - Modelare



Masurarea tendintei centrale



Tipuri de scale	Indicatori ai tendintei centrale			
	Modulul	Mediana	Media aritmetica	Media geometrica
Nominala	X	-	-	-
Ordinala	X	X	-	-
Interval	X	X	X	X
Proportionala	X	X	X	X





Masurarea tendintei centrale

■ Grupul modal (modulul)

- Grupul care cuprinde cele mai multe componente comparativ cu celelalte grupuri

■ Mediana

- Valoarea care imparte numarul de observatii in doua grupuri egale

■ Media aritmetica

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

■ Media geometrica

$$X_G = \sqrt[n]{\prod_{i=1}^n X_i}$$

$$\bar{x}_p = \frac{\sum_{i=1}^n p_i x_i}{\sum_{i=1}^n p_i}$$





Analiza variatiei

- Distributia frecventelor (absolute si relative)
- Amplitudinea variatiei ($X_{\max} - X_{\min}$)
- Varianta
$$\sigma^2 = \sum_{i=1}^n \frac{(x_i - \bar{x})^2}{n-1}$$
- Abaterea medie patratica (standard)
- Coeficientul de variatie (variatiia relativa)

$$CV = \frac{\sigma}{\bar{x}}$$





Distributia frecventelor

■ Perceptia gustului pentru berea Redd's

Eticheta	Codificare	Frecventa	Frecventa (%)	Procent valid	Frecvente cumulate
Cel mai neplacut	1	0	0,0	0,0	0,0
Foarte neplacut	2	2	6,7	6,9	6,9
Neplacut	3	6	20,0	20,7	27,6
Nici/Nici	4	6	20,0	20,7	48,3
Placut	5	3	10,0	10,3	58,6
Foarte placut	6	8	26,7	27,6	86,2
Cel mai placut	7	4	13,3	13,8	100,0
Valori lipsa	9	1	3,3		
Total		30	100	100	



Analiza variatiei

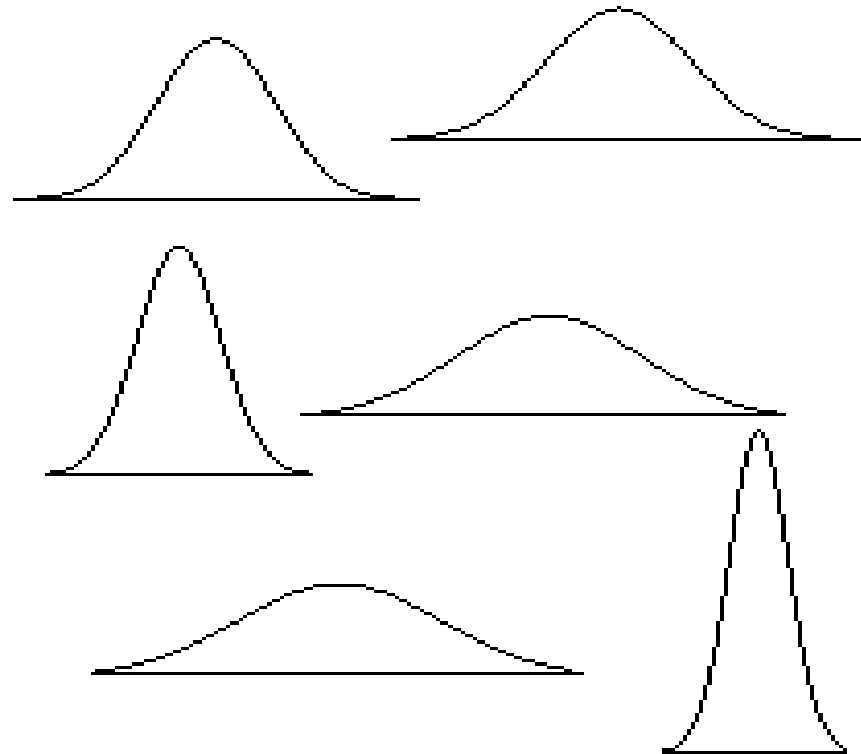


Indicatori ai dispersiei	Tipuri de scale			
	Nominale	Ordinale	Interval	Proportionale
Frecvente	X	X	X	X
Amplitudine	-	X	X	X
Coeficient de variatie	-	-	X	X
Varianta	-	-	X	X
Abaterea standard	-	-	X	X



Distributia normala

- Este constituita dintr-o familie de distributii care au reprezentari grafice asemanatoare unui clopot;





Distributia normala

- Este importanta pentru ca **majoritatea instrumentelor statistice utilizate in analiza primara au ca premisa existenta unei distributii normale** (ex.: *testul Student, Mann-Whitney, ANOVA, Pearson, regresia, nivelarea exponentiala Brown, etc.*).
- **Ipoteza distributiei normale:**
 - Pentru δ ales, δ procente dintre valorile inregistrate ale variabilei se vor afla in intervalul:

$$(\bar{x} - \sigma t_{\delta}; \bar{x} + \sigma t_{\delta})$$





Distributia normala

- **Aplatizarea**: reprezinta o masura a inaltimii relative a “clopotului” definit de distributia frecventelor variabilei.

$$K = \frac{n \sum_{i=1}^n (x_i - \bar{x})^4}{\left(\sum_{i=1}^n (x_i - \bar{x})^2 \right)^2} - 3$$

- **Asimetria**: tendinta variatiilor valorilor observate fata de medie de a fi mai mari catre unul dintre capetele intervalului de valori.



$$G = \frac{\sqrt{n} (n-1)}{n-2} \times \frac{\sqrt{n} \sum_{i=1}^n (x_i - \bar{x})^3}{\left(\sum_{i=1}^n (x_i - \bar{x})^2 \right)^{3/2}}$$



Distributia normala

■ **Inaltimea curbei** este data de formula:

$$h = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{2\sigma^2}}$$

■ **Testarea normalitatii distributiei:**

- variabile parametrice: testul Kolmogorov-Smirnov
- variabile ordinale: testul Shapiro-Wilk
- variabile nominale: testul χ^2





Testul Kolmogorov-Smirnov

- Utilizat pentru estimarea normalitatii distributiei acolo unde se poate (are sens) calcula media si abaterea medie patratica.
- De fapt, determina care dintre ipoteze va fi adoptata:
 - ➡ H_0 : NU exista diferente semnificative statistic intre distributia variabilei investigate si distributia normala.
 - ➡ H_1 : Exista diferente semnificative statistic intre distributia variabilei investigate si distributia normala.





Testul Kolmogorov-Smirnov

■ Se calculeaza utilizand:

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n I_{(x_i < \bar{x})}$$

$$K_s = \max_{i=1}^n \left| F(x_i) - \frac{i-1}{n} \right|; \left| \frac{i}{n} - F(x_i) \right|$$





Testul Kolmogorov-Smirnov

■ H_0 este acceptata daca:

$$K_s \left(\sqrt{n} + 0,12 + \frac{0,11}{\sqrt{n}} \right) \leq c_\alpha$$

unde:

α	0,85	0,9	0,95	0,975	0,99
c_α	1,138	1,224	1,358	1,480	1,628





Testul Shapiro-Wilk

- Utilizabil **atat pentru variabile non-parametrice, cat si pentru variabile parametrice** (aici insa testul Kolmogorov-Smirnov este mai puternic).
- Determina daca va fi adoptata ipoteza nula sau ipoteza alternativa:
 - ➡ H_0 : **NU exista diferente semnificative statistic intre distributia variabilei investigate si distributia normala.**
 - ➡ H_1 : **Exista diferente semnificative statistic intre distributia variabilei investigate si distributia normala.**





Testul Shapiro-Wilk

- Se calculeaza utilizand:

$$W = \frac{\left\{ \sum_{i=1}^n a_i x_i \right\}}{\sum_{i=1}^n (x_i - \mu)^2}$$

- a_i reprezina parametrii dati, obtinuti din tabele statistice pe baza medianei, numarului de valori (categorii) distincte si dimensiunii esantionului investigat.
- $p(W) > 0,05 \Rightarrow H_0$ este acceptata
altfel $\Rightarrow H_1$ este acceptata



