



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

- Analiza asociativa -



Analiza asociativa a datelor

- Presupune măsurarea gradului de asociere sistematică a valorilor a două variabile sub aspectul:

- **Directiei** (*directa / inversa*)

- **Intensitatii** (forteii)

- puternica;
- moderata;
- slaba;
- foarte slaba;
- inexistentă.

- **Semnificatiei statistice.**



Analiza asociativa a datelor



- 4 tipuri de asocieri statistice:
 - ➡ **Non-monotone** (prezenta sau absenta unei variabile este sistematic asociata cu prezenta sau absenta valorilor unei alte variabile) – *nu permite identificarea clara a directiei* (ex.: utilizatorii PlayStation sunt copii, nu adulti);
 - ➡ **Monotone** (se poate identifica directia asociatiilor, dar nu si intensitatea) – ex.: inaltimea unei persoane influenteaza marimea pantofilor);
 - ➡ **Liniara si non-liniara** (se pot identifica directia si intensitatea)
 - ➡ daca se cunoaste valoarea unei variabile poate fi identificata automat valoarea celeilalte variabile;



Analiza asociativa a datelor



- Indicatori utilizati:
 - Coeficientul de corelatie r_{phi} (Φ);
 - Coeficientul de contingenta C ;
 - Coeficientii de corelatie a rangurilor ρ (Spearman) si γ (Kruskal si Goodman);
 - Coeficientul de corelatie r (Pearson);



Coeficientul de corelatie r_{phi}



- Utilizat pentru identificarea asocierilor existente intre doua variabile dihotomice.
- Coeficientul de corelatie:

$$r_{\text{phi}} = \frac{ad - bc}{[(a + b) (c + d) (a + c) (b + d)]^{1/2}}$$

- **Directia asocierii** este data de semnul lui r_{phi}
- **Coeficientul de determinare** $(r_{\text{phi}})^2$ determina gradul in care variabila dependenta este influentata de variabila independenta.
- **Gradul de semnificatie al asocierii** dintre opiniile subiectilor constituiti in cele doua esantioane este determinat cu ajutorul testului χ^2 in varianta *Fisher* (esantioane independente) sau *McNemar* (esantioane dependente).



Coeficientul de corelatie r_{phi}



➡ Intensitatea corelatiei:

- ➡ $0 \geq |r_{\text{phi}}| \geq 0,2 \Rightarrow$ corelatie **nesemnificativa**;
- ➡ $0,2 > |r_{\text{phi}}| \geq 0,4 \Rightarrow$ corelatie **foarte slaba**;
- ➡ $0,4 > |r_{\text{phi}}| \geq 0,6 \Rightarrow$ corelatie **slaba**;
- ➡ $0,6 > |r_{\text{phi}}| \geq 0,8 \Rightarrow$ corelatie **moderata**;
- ➡ $0,8 > |r_{\text{phi}}| \geq 1 \Rightarrow$ corelatie **puternica**;





Coeficientul de contingenta C

- Utilizat pentru variabile categoriale (tabele de contingenta de dimensiunea $r \times k$);
- Dependent de conditiile de calcul ale χ^2 .
- Nu poate fi determinata **directia** asocierii!
- **Coeficientul de contingenta:**

$$C = \sqrt{\frac{X_c^2}{N + X_c^2}}$$

N.B.: Coeficientii de contingenta nu pot fi comparati decat daca provin din tabele de contingenta de aceeasi dimensiune!





Coeficientul de contingenta C

- Intensitatea corelatiei:

$$C \in \left[0, \sqrt{\frac{k-1}{k}} \right)$$

- Intervalul astfel obtinut este impartit in 5 sub-intervale egale
- Nu exista un test de semnificatie al Coeficientului de contingenta!





Coeficientul de corelatie Spearman

- Denumit si “coeficientul de corelatie al rangurilor”, este utilizat doar pentru a compara variabile masurate pe scala ordinala sau interval.

- Coeficientul Spearman de corelatie a rangurilor

$$\rho = 1 - \frac{6 \sum_{i=1}^n D_i^2}{n (n^2 - 1)}$$

- D_i reprezinta diferentele dintre rangurile unei inregistrari (respondent);





Coeficientul de corelatie Spearman

- ➡ **Directia asocierii** este data de semnul lui ρ .
- ➡ **Intensitatea asocierii** este data de valoarea lui ρ .
- ➡ **Gradul de semnificatie** al asocierii este determinat cu ajutorul testului de semnificati al coeficientului Spearman, calculat cu ajutorul formulei:

$$Z_c = \frac{\rho}{\frac{1}{\sqrt{n-1}}}$$





Coeficientul de corelatie Spearman

- Bazat pe ipoteza nula:
 - ➡ H_0 : coeficientul de corelatie NU DIFERA in mod semnificativ de zero.
 - ➡ H_1 : coeficientul de corelatie DIFERA in mod semnificativ de zero.
- Valoarea calculata z_c a testului se compara cu valoarea tabelata z_t a acestuia, obtinuta in functie de probabilitatea de garantare a rezultatului.
 - ➡ $-z_t \leq z_c \leq z_t$: se accepta ipoteza nula
 - ➡ altfel: se accepta ipoteza alternativa



Coeficientul de corelatie γ (Goodman si Kruskal)



- Se poate utiliza atunci cand coeficientul Spearman nu ofera valori concludente.
- Valoarea lui γ , la fel ca si ρ , poate lua valori in intervalul -1 si +1.

$$\gamma = \frac{P - Q}{P + Q}$$

- P se obtine inmultind valoarea frecventei din coltul din stanga sus a tabelului de frecvente asociat cu valorile de pe randul urmator, fara cea care se afla imediat sub ea, si cu celelalte valori ale tabelului, dupa care se insumeaza cu produsul dintre prima valoare a randului urmator si suma frecventelor incepand de pe randul urmator, insa din nou fara valoarea aflata imediat sub ea, iterativ. Q se calculeaza dupa aceleasi reguli, insa incepand din dreapta sus.



Coeficientul de corelatie γ (Goodman si Kruskal)



Categorie de varsta	Intentiile de cumparare				
	1	2	3	4	5
1 – (sub 20 de ani)	5	10	15	20	25
2 – (20 – 40 de ani)	10	10	15	20	20
3 – (40 – 60 de ani)	20	15	15	10	5
4 – (peste 60 de ani)	35	25	20	5	1

$$\begin{aligned}
 P = & 5 (10+15+20+20+20+15+15+10+5+25+20+5+1) + \\
 & 10 (15+20+20+15+10+5+20+5+1) + 15 (20+20+10+5+5+1) + \\
 & 20 (20+5+1) + 10 (15+15+10+5+25+20+5+1) + \\
 & 10 (15+10+5+20+5+1) + 15 (10+5+5+1) + 20 (5+1) + \\
 & 20 (25+20+5+1) + 15 (20+5+1) + 15 (5+1) + 10 \times 1 = 6815
 \end{aligned}$$





Coeficientul de corelatie Pearson

- ➔ Utilizat atunci cand cel puțin una dintre variabile este **metrica**, variabilele sunt continue, au distributii normale si dispersii asemanatoare.
- ➔ Calculat cu ajutorul formulei:

$$r = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{\sqrt{\left(n \sum_{i=1}^n (x_i)^2 - \left(\sum_{i=1}^n x_i \right)^2 \right) \left(n \sum_{i=1}^n (y_i)^2 - \left(\sum_{i=1}^n y_i \right)^2 \right)}}$$

- ➔ **Directia asocierii** este data de semnul lui r.
- ➔ **Intensitatea asocierii** este data de valoarea (absoluta) a lui r.
- ➔ **Coeficientul de determinare r^2** indica gradul in care variatiile de la nivelul variabilei independenta explica variatiile constatate la nivelul variabilei dependente.





Coeficientul de corelatie Pearson

- **Gradul de semnificatie al asocierii** este stabilit prin testarea semnificatiei statistice a coeficientului r .
- Bazat pe ipoteza nula:
 - ▶ H_0 : nu exista o corelatie intre variabilele investigate (r nu este in mod semnificativ diferit de zero).

$$Z_c = \frac{1,1513 + \log_{10} \left(\frac{1+r}{1-r} \right)}{1/\sqrt{n-3}}$$

- Valoarea calculata Z_c a testului se compara cu valoarea tabelata Z_t a acestuia, obtinuta in functie de probabilitatea de garantare a rezultatului.
 - ▶ $-Z_t^2 \leq Z_c^2 \leq Z_t^2$: se accepta ipoteza nula
 - ▶ altfel: se respinge ipoteza nula

Analiza asociativa a datelor



- **Indicatori utilizati:**

- **Variabile categoriale:**

- variabile dihotomice => r_{ϕ} ;
- variabile ordinale sau interval => ρ sau γ ;
- variabile nominale, non-dihotomice: C .

- **Variabile proportionale => r ;**





Corelatii partiale

- Nu intotdeauna o corelatie observata reprezinta asocieri directe intre variabile (**corelatia nu implica cauzalitatea**)!
- **Exemplu:** a fost demonstrata o relatie directa si puternica intre consumul de inghetata si infractiunile saptamanale (numarul de infractiuni saptamanale) din New York!! (sa le propunem sa interzica inghetata!?!)

– **Motivul:** ambele sunt influentate de temperatura!



Corelatii partiale



- **Coeficientul de corelatie partiala** masoara gradul de asociere dintre doua variabile, excluzand in prealabil efectele asupra variabilei dependente a una sau mai multe variabile de control (independente)!

- **Exemple:**

- efectul asupra cresterii vanzarilor al unui spot promotional, atunci cand este exclus efectul reducerilor de pret asociate;
- perceptia consumatorilor asupra calitatii produsului este influentata de perceptia asupra pretului, in conditiile in care se exclude perceptia asupra imaginii marcii, etc.





Corelatii partiale

- ➡ **Coeficientul de corelatie partiala** dintre variabila independenta X si variabila dependenta Y, in conditiile in care este exclus (“controlat”) efectul variabilei independente (de contro) Z, notat r_{xy-z} .
- ➡ Calculat cu ajutorul formulei:

$$r_{xy-z} = \frac{r_{xy} - (r_{xz}) (r_{yz})}{\sqrt{1 - (r_{xz}^2)} \sqrt{1 - (r_{yz}^2)}}$$

- ➡ unde: r_{xy} este coeficientul de corelatie dintre variabilele X si Y, etc.





Corelatii partiale

- Coeficientii de corelatie partiala au o **ordine** asociata, care indica numarul de variabile de control utilizate (coeficientii de corelatie directi sunt “de ordin zero”).
- Formula de calcul este recursiva. Coeficientul de corelatie partiala de ordin n (cu n variabile de control $Z=\{Z_1, Z_2, \dots, Z_n\}$) se calculeaza pe baza a trei coeficienti de corelatie partiala de ordin n-1:

$$\rho_{xy-z} = \frac{\rho_{xy-Z \setminus \{z_0\}} - \rho_{xz_0-Z \setminus \{z_0\}} \rho_{yz_0-Z \setminus \{z_0\}}}{\sqrt{1 - \rho_{xz_0-Z \setminus \{z_0\}}^2} \sqrt{1 - \rho_{yz_0-Z \setminus \{z_0\}}^2}}$$



Corelatii partiale



- ➡ **Directia asocierii** este data de semnul lui r_{xy-z} .
- ➡ **Intensitatea asocierii** este data de valoarea (absoluta) a lui r_{xy-z} .
- ➡ Coeficientii de corelatie partiala sunt utilizati pentru o predictie mai exacta a directiiei asocierilor (cauzalitatii), dar nici ei nu trebuie considerati exclusivi!



Corelatii partiale



Atitudinea fata de berea Redd's.

<i>Respondent</i>	<i>Atitudinea (scala Stapel)</i>	<i>Ore petrecute lunar in cluburi</i>	<i>Importanta relativa a pretului</i>
<i>1</i>	6	10	3
<i>2</i>	9	12	10
<i>3</i>	8	12	4
<i>4</i>	3	4	1
<i>5</i>	10	12	10
<i>6</i>	4	6	1
<i>7</i>	5	8	7
<i>8</i>	2	2	4
<i>9</i>	10	18	8
<i>10</i>	9	9	10
<i>11</i>	10	17	8
<i>12</i>	2	2	5





Corelatii partiale

➡ Coeficientii simpli (de ordin zero) intre variabile sunt:

➡ $r_{x_1y} = 0,9361$

➡ $r_{yx_2} = 0,7334$

➡ $r_{x_1x_2} = 0,5459$

$$r_{x_1y-x_2} = \frac{r_{xy} - (r_{xz}) (r_{yz})}{\sqrt{1 - (r_{xz}^2)} \sqrt{1 - (r_{yz}^2)}}$$

$$r_{x_1y-x_2} = \frac{0,9361 - (0,5495) (0,7334)}{\sqrt{1 - (0,5495^2)} \sqrt{1 - (0,7334^2)}} = 0,9386$$



Corelatii partiale



- In conditiile in care $r_{xy-z} = 0,9386$ se poate concluda ca preferinta pentru berea Redd's este foarte puternic asociata cu numarul de ore petrecut de respondenti in cluburi, chiar si atunci cand este exclus efectul perceptiei respondentilor asupra *pretului* berii.



