

**APLICAREA METODEI STATISTICE DE ANALIZĂ MULTIVARIATĂ
„ANALIZA ÎN COMPONENTE PRINCIPALE” ÎN IDENTIFICAREA
SIMILARITĂȚII MĂRCILOR AUTOHTONE DE „BERE” CU CELE
DE PE PIAȚA UE**

Violeta GOLEA
Cercetător științific stagiar, IEFS

In the article is outlined the essence and importance of multivariate statistical method "Principal Components Analysis", and its application in identifying similarity between national "beer" product and the EU "beer" products. ACP technique, use mathematical procedures to convert a set of related observations into a set of values of uncorrelated variables called "principal linear components". Usually the number of principal components is less or equal to the original variables. This transformation is defined in such way, that first component have the biggest variation. The main components are guaranteed to be independent only if the data set follows a normal distribution

Desfășurarea oricărei activități umane este formată și permanent renovată de apariția de noi date și informație care este utilizată pentru descoperirea și implementarea de noi metode, cunoștințe asupra fenomenelor socio-economice.

O dată cu trecerea timpului în urma efectuării a multor cercetări în domeniul industriei băuturilor alcoolice, reapare pofta de noi cercetări și implementări a evoluției calității produselor existente și perfecționarea lor pentru a fi într-un pas cu brandurile europene, de aceea informația apărută și expusă tinde să răspundă la multe întrebări.

Deci statistica multivariată oferă o gamă largă de metode în analiza unei dimensiuni mari de date. Cea mai des folosită și aplicată în analiza și interpretarea datelor multivariate este metoda „Analiza în Componente Principale” (ACP). Această metodă a fost inventată în anul 1901 de către vestitul matematician englez Karl Pearson, care a fost creditat pentru stabilirea disciplinei statistica matematicii.[1:1]

Această metodă factorială a apărut pentru a rezolva probleme din cadrul mai multor domenii ca:

- Reducerea complexității datelor, unde poate fi înlocuit un masiv de date de dimensiuni mari prin masive de dimensiuni mai mici .
- Evidențierea și fixarea paternului corelațiilor dintre variabile.
- Determinarea variabilelor latente care se află în spatele variabilelor măsurate – de exemplu problema similară descoperirii celor care mănuiesc păpușile într-un teatru: comportarea, varianța variabilelor măsurate poate fi regăsită din varianța unor variabile ascunse, care le determină prin asociere.

ACP este foarte des utilizată în cercetările economice, precum și în cadrul altor domenii de activitate. Atunci când cercetătorii se confruntă cu o mare complexitate a proceselor economico -socio-financiare și este nevoie de analizat un număr mare de variabile, care deseori se dovedește a fi problematică din cauza numărului mare de date și de greutatea de-a efectua calcule asupra lor. În plus există o șansă destul de mare ca datele inițiale să fie foarte bine corelate. Deci cercetătorilor le este necesar un instrument puternic pentru a efectua analiza și a putea rezolva astfel de probleme, de-a efectua o analiză corectă și concludentă.

ACP, folosește procedurile matematice pentru a converti un set de observații corelate într-un set de valori ale variabilelor liniare necorelate numite componente principale. De obicei numărul de componente principale este mai mic sau egal cu numărul de variabile originale. Această transformare este definită în așa fel încât prima

componentă să posede cea mai mare variație. Componentele principale sunt garantate a fi independente numai în cazul în care setul de date urmează o distribuție normală. Deci distribuția normală multivariată în teoria probabilității și statisticii este o generalizare de o dimensiune (univariată) sau reprezintă o distribuție normală de dimensiuni mari.

Importanța acestei metode este faptul că produce cea mai bună bază, astfel în sensul minimizării erorii. Coeficienții din teoremă sunt variabile aleatoare și baza de extindere depinde de proces. De fapt, funcțiile de bază ortogonale utilizate în această metodă sunt determinate de funcția covarianță a procesului. Se spune că metoda ACP se adaptează la proces pentru a produce cea mai bună bază posibilă.

La momentul actual ACP este folosită ca instrument de explorare în analiza datelor, care se ocupă cu analiza seturilor de date pentru a rezuma principalele lor caracteristici într-o formă mai înțeleasă, utilizând graficele vizuale.

Prima componentă principală extrasă este acea combinație liniară de variabile care preia maximul posibil din varianța datelor inițiale. Ce-a de-a doua componentă principală preia mai puțină varianță și așa mai departe. Dacă primele câteva componente principale preiau 80% sau mai mult din varianța datelor inițiale, atunci scopul reducerii dimensionalității va fi atins [1:3].

Calculul este făcut prin algoritmi ai algebrei liniare, aplicând fie matricea dreptunghiulară a datelor originale, fie matricea pătratică a coeficienților de corelație. În teoria matematică ce fundamentează acești algoritmi se folosește termenul de valoare proprie a unei matrice. Valorile proprii măsoară cantitatea de varianță explicată de fiecare componentă principală. Ele descresc o dată cu indexul componentei, prima componentă principală având valoarea proprie maximă. Pe de altă parte, suma valorii proprii este egală cu p (numărul variabilelor inițiale). De obicei sunt reținute doar componentele principale ce au valorile proprii mai mari decât 1. Valorile proprii exprimă importanța componentelor principale.

Această metodă este caracterizată de unele puncte forte și slabe, care au fost identificate de către cercetători pe parcursul evoluției sale: [1:2]

Avantajele utilizării ACP sunt:

- Atât atributele obiective cât și cele subiective pot fi folosite cu condiția ca atributele subiective să fie convertite în scoruri.
- Analiza factor poate fi utilizată pentru a identifica dimensiunile ascunse sau construcțiile care nu pot reieși din analiza directă.
- Este ușor și necesită finanțe puține pentru a o elabora.

Dezavantajele acestei metode:

- Utilitatea depinde de capacitatea cercetătorilor de a colecta un set suficient de atribute ce se referă la produs. Dacă atributele sunt important tratate atunci valoarea procedurii este redusă.
- În cazul în care seturi de variabile observate sunt foarte asemănătoare între ele și distincte de alte elemente, ACP va atribui un singur factor pentru ei. Acest lucru poate face mai greu de a identifica factorii care surprind relațiile mult mai interesante.
- Denumirea de factori ar putea necesita cunoștințe de fundal sau teorie, deoarece mai multe atribute pot fi extrem de corelate pentru un motiv neimportant. [1:2].

Se constată că la moment în țară utilizarea metodei ACP este mai puțin abordată din punct de vedere practic cât și teoretic. Mai des utilizat este în cazul unor cerințe despre acumularea informației despre un produs sau obiect în dependență de regiune, localitate la cerința unor firme private cu caracter confidențial.

La moment un produs nu poate fi cercetat independent de evoluția peții economice mondiale. Aceasta este oportun doar în cazul unei țări mici cu o economie de

piață mică. Însă pentru o dezvoltare durabilă o astfel de economie necesită o deschidere mare către piața mondială.

Aplicarea metodei Analiza în Componente Principale și interpretarea rezultatelor obținute

Evaluarea calitativă a formei de repartizare a datelor statistice obținute s-a efectuat în baza folosirii software-ului SPSS (Statistical Package for the Social Sciences), este unul dintre cele mai accesibile soft, folosit de organizații comerciale, guvernamentale și academice pentru a rezolva probleme legate de afaceri etc.

Scopul aplicației este acela de a identifica similaritatea berii, produse de producătorii autohtoni cu brendurile expuse pe piața UE de către producătorii străini utilizând metoda de statistică multivariată Analiza în Componente Principale, pentru care s-a selectat opt caracteristici principale organo-leptice de caracterizare a berii și 17 denumiri de bere din diferite țări ale UE.

În rezultatul calculării a fost obținută o matrice de corelație unde se caracterizează interacțiunea factorilor și caracteristicile prezente.

În baza tabelului de corelației și a tehnicii de prelucrare pas cu pas s-a depistat că din cele 8 caracteristici inițiale doar șase din ele au o corelație puternică între ele explicând 74% din varianța generală a norului de puncte. Aceste caracteristici sunt: persistența spumei, înălțimea spumei, transparența, culoarea, aciditatea, alcoolul etilic. Aceste concluzii sunt argumentate și de calcularea testului KMO and Bartlett's=0,721 și testul Chi-patrat=44,9 la un nivel de semnificație 0,01 unde se acceptă ipoteza alternativă, adică variabilele sunt corelate între ele și metoda ACP are sens.

Tabelul 1. Testul KMO și Bartlett's

KMO and Bartlett's Test

	Kaiser-Meyer-Olkin Measure of Sampling Adequacy.	,721
Bartlett's Test of Sphericity	Approx. Chi-Square	44,944
	df	15
	Sig.	,000

Deci o dată cu aplicarea acestei metode s-a obținut două axe factoriale care explică varianța caracteristicilor de bază cu aproximativ 74%.

Tabelul 2. Varianța componentelor selectate

Varianța totală explicată

Component	Initial Eigenvalues		
	Total	% of Variance	Cumulative %
1	3,351	55,852	
2	1,053	17,550	
3	,853	14,223	87,625
4	,406	6,764	94,390
5	,206	3,426	97,816
6	,131	2,184	100,000

Conform statisticilor din tabelul nr. 3. care prezintă încărcările factorilor, se constată că componenta unu este definită de variabilele *persistența spumei și culoare*, iar componenta doi este definită de *transparența și prezența alcoolului etilic*.

Tabelul 3. Matricea Componentelor

Component Matrix^a

	Component	
	1	2
persist_spumei	,865	,339
inal_spumei	,766	-,311
transparenta	,733	,477
culoare	,915	-,175
aciditatea	,462	,397
alcool_etilic	,655	-,653

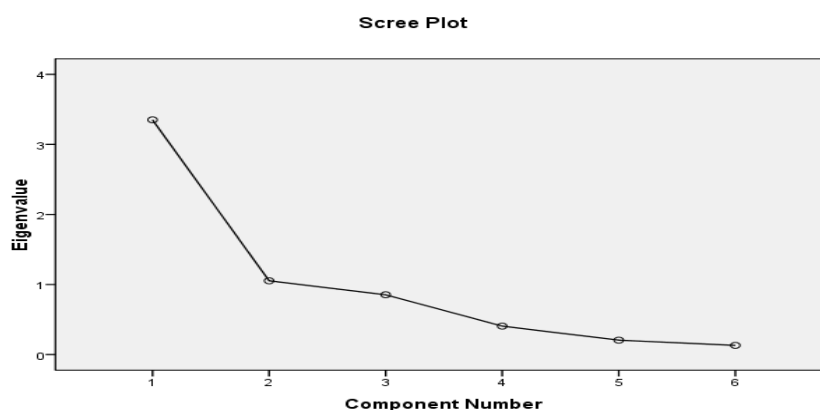


Fig. 1. Reprezentarea variabilelor corelate

Concluzie

Analizând simultan cele două grafice de mai jos putem ușor depista ce caracteristici forte și slabe deține fiecare dintre tipurile de bere analizat, precum și identificarea similarităților cât și a diferențelor dintre produse.

Cea mai bună bere care se încadrează în standardele europene la caracteristica *înălțimea spumei* este marca de bere **vitanta** care este un produs autohton și concomitent corespunde standardelor internaționale de merceologie iar la variabila persistența culorii fiind similară cu berea germană Neumarkt. O remarcă este asupra berii bulgărești boliarka care dispune de prezența unei acidități peste măsură de nivelul teoretic, în favoarea calității înălțimea spumei.

Analizând situația după cea de-a doua componentă *transparența și prezența alcoolului etilic*, se evidențiază o situație bună asupra berii ciuc, albacher, burgher care sunt similare între ele după o concentrare de alcool destul de mare și există între ele o legătură de corelație puternică, în schimb la produsele de bere: orașul vechi, noroc, bergangrier, chișinău există o similaritate puternică după variabila *transparență* și o disfuncție asupra concentrării mici de alcool etilic. Deci putem spune că atât berea autohtonă cât și cea de pe piața UE sunt similare în respectarea unor standarde ca componenta alcoolului etilic, *transparență*, *persistența spumei*.

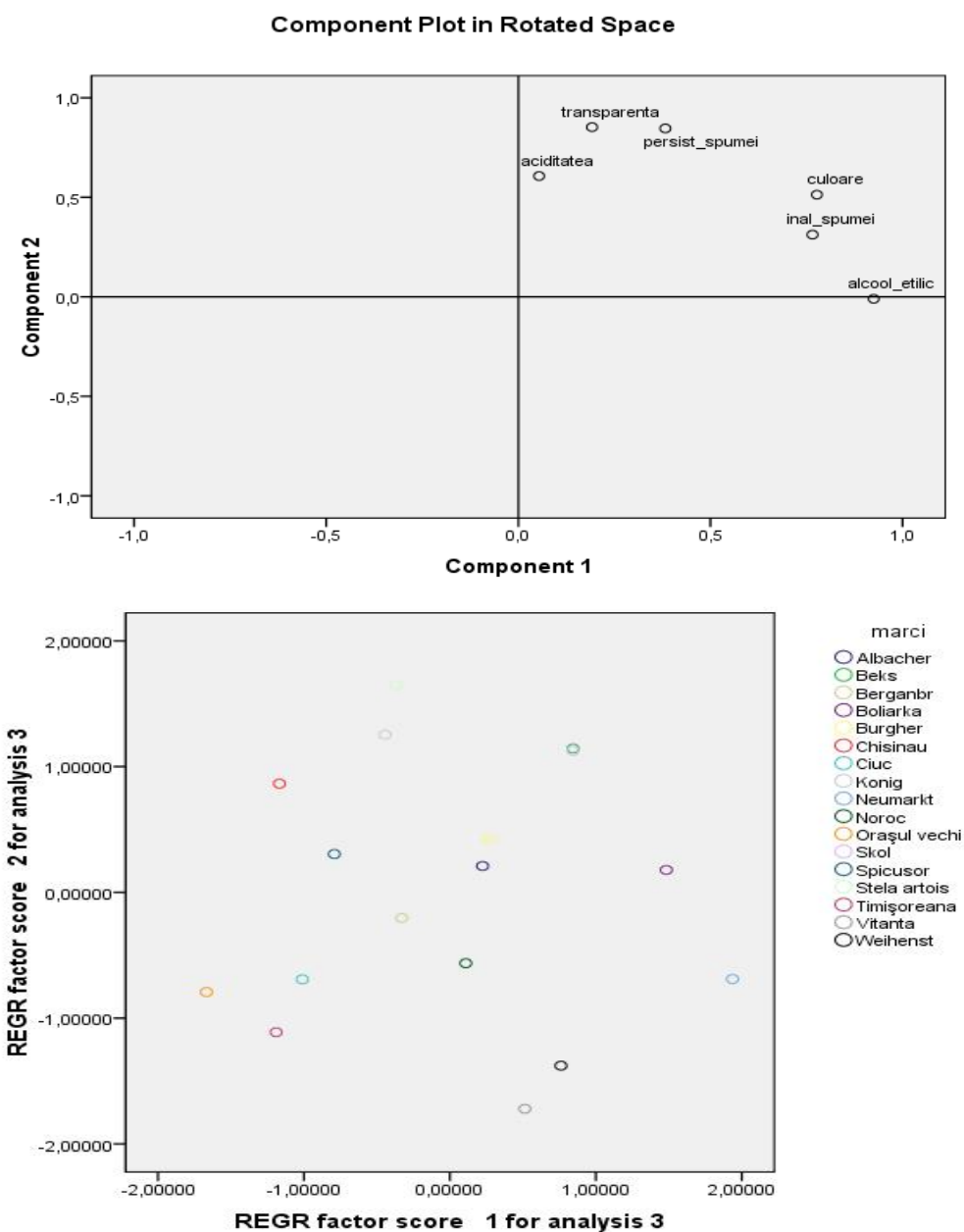


Fig.2. Repartizarea mărcilor de bere în spațiul componentelor principale

Recomandări:

- Ar fi necesar ca marketologii autohtoni să țină cont de calitatea produselor bune pentru a le putea promova pe piața UE.
- Implementarea metodelor avansate de obținere a unor produse calitative pentru a stimula exportul de băuturi alcoolice autohtone și a readuce venituri în țara noastră.
- Atragerea investitorilor străini pentru încheierea unor contracte de vânzare-cumpărare a băuturilor alcoolice.

Referințe bibliografice

1. Spîrcu L., Analiza datelor, Aplicații economice, ed. ASE, București 2005.
2. Spîrcu L., Pașilea V., Modele de scoring în studii și cercetări de calcul economic și cibernetică economică, 1995.
3. Spîrcu-Tiberiu, cursuri de doctorat 2007.