

METODA DISTANȚEI MEDII.

Analiză statistică a compozițiilor elementale

dr. Eugen S. Teodor

În cadrul programului Arheomet se implementează cooperarea specialiștilor în istorie și arheologie cu inginerii de la IFIN, primii selectând un set de obiecte pentru analiză, iar cei din urmă furnizând rezultatele testelor de laborator. Aceste rezultate cuprind, alături de identificarea fiecărei piese (nume, număr de inventar, dar, mai ales, un număr de identificare în baza de date, așa numitul ID) un set de date numerice, reprezentând prezența fiecărui element chimic în structura metalului. Funcție de tipul generic de metal (numit, în limbaj muzeografic, „aur”, „argint”, „bronz”), dar și de metoda folosită, tabelul de date rezultat poate avea 5, dar și 15 coloane de cifre, respectiv elemente chimice regăsite. Aceste date pot oferi importante indicii asupra sursei de materie primă sau material reciclat. Doar că datele brute, respectiv rezultatele măsurătorilor elementale, nu înseamnă, în sine, nimic. Aceste date trebuie interpretate. Acest lucru îl poate face fizicianul – sau orice altcineva care poate realiza o analiză statistică (care este, în acest caz, întâi o analiză de grupare, și abia apoi una statistică). Mult mai normal este însă ca interpretarea datelor să fie făcută de istoric/ arheolog, pe un model deja consacrat al cooperării dintre arheologie și arheometrie, fiindcă adesea soluțiile finale sunt generate de recursul la obiect, prin confruntarea datelor de măsurătoare cu unele proprietăți „muzeografice” ale obiectului.

În general, seriile scurte de măsurători (până la 20 de piese), mai ales cu un set scurt de elemente chimice identificate, nu presupun analize dificile, nefiind necesar un efort metodologic deosebit. Nu la fel stau lucrurile atunci când se confruntă serii mari de piese, așa cum s-a întâmplat în derularea fazei II a proiectului (iulie-septembrie 2006), când a apărut spre analiză un lot de aproape 180 de monede medievale de argint, lot pe care apăreau 14 variabile chimice. De la acest moment devine obligatoriu aportul unei aplicații informatice dedicate, pentru a sorta și grupa această masă semnificativă de date. Acest lot de monede (vezi [Anexa 1](#)) a făcut obiectul unei aplicații-test, care va fi expusă mai jos.

Metoda de analiză, numită *metoda distanței medii* (Average Gap) studiază „distanțele” dintre obiectele analizei, comparând fiecare obiect în parte cu toate celelalte obiecte, pentru oricare număr de variabile, clasând relația de diferențiere funcție de niște „plafoane” prestabilite (stabilite experimental, în fapt). Procedural, primul obiect al seriei devine „obiect-standard”, caracteristicile sale devenind standardul de comparație. Pentru fiecare dintre variabile se stabilește maximul (M) și minimul (m) seriei, diferența dintre ele (D) fiind împărțită la numărul de obiecte ale seriei (N), rezultând „pasul” specific fiecărei variabile (P). Fraza de mai sus se comprimă în expresia foarte simplă: $P = D : N$.

La momentul următor (în sens logic, fiindcă procesul este instantaneu) se calculează diferențele constatate între „obiectul-standard” (obiectul de referință) și toate „obiectele-țintă” (toate celelalte obiecte din lot), având ca măsură „pasul” (P). Notăm mai jos cu S – obiectul standard și cu T – obiectul-țintă, cu care facem comparație, cu $V_1, 2, \dots$ – numărul variabilei, cu nV – numărul de variabile:

$$X = [(SV_1-TV_1) + (SV_2-TV_2) + \dots + (SV_n-TV_n)] : nV,$$

unde X reprezintă diferența *medie* a variabilelor comparate; valoarea lui X este apoi convertită într-o clasificare, având următoarele valori:

$$0 = X < 0,1$$

$$1 = X [0,1...0,5]$$

$$2 = X [0,5...1,5]$$

$$3 = X [1,5...3]$$

$$4 = X [3...5]$$

$$5 = X > 5$$

și în care valoarea 1 = Pm (pasul mediu al tuturor variabilelor)

Rezultatele comparațiilor se transcriu automat într-o tabelă nouă, în care se transferă numărul de identitate a obiectul-standard, a obiectului țintă, și „clasarea” relației dintre ele (numărul de mai sus). După depășirea acestei faze, procesul este reluat cu următorul obiect-țintă (concret, a treia înregistrare din tabela cu date brute), apoi la următorul, până la epuizarea listei (în cazul nostru 180). Aici se termină prima buclă majoră în execuția programului. Execuția continuă cu alegerea următorului obiect-standard (al doilea din listă, deci), care urmează a fi comparat cu toate celelalte (cu excepția primului, care nu mai este vizitat). Pentru fiecare operațiune descrisă se transcrie rezultatul în tabela de clasificări; la lotul de 188 de monede (inițial; ulterior câteva au fost excluse din analiză), în tabela nou se făcuseră, automat, 17954 de înregistrări... Pentru că, nu-i așa, sunt permutări de 188 obiecte (nu faceți socoteala; sunt exact 376 de înregistrări „parazite”, cu funcție strict tehnică).

Ce facem acum cu o tabelă cu 18000 de înregistrări?... Majoritatea acestora sunt clasări cu valoare 5 (cu semnificația „neanalog”), care ies din discuție; rămân 323 clasări cu valoarea 4, 96 clasări cu valoarea 3 și 9 clasări cu valoarea 2. Clasările cu valoare 4 au valoare pur orientativă, semnalând mai degrabă o linie de înrudire colaterală, și sunt utile în judecarea seriilor scurte de obiecte comparate.

Valoarea de colaterală a obiectelor care au relație de tip 4 a fost testată pe loturi de piese anterioare, în care au rezultat relații de genul:

ID1~2 și ID3~ID4 – la nivel analogic 2 (variație de sub 1,5%)

dar

ID1~ID3 și ID2~ID4 – la nivel analogic 4 (variație de la 3 la 5%)

Relațiile de mai sus sunt foarte explicite și demonstrează la nivel 2 derivarea nemijlocită, iar la 4 – doar o derivare colaterală; cu alte cuvinte, nivelul analogic 4 nu este util pentru stabilirea componenței grupelor, ci pentru a surprinde *înrudirea dintre grupe*, într-o etapă de prelucrare avansată a datelor.

Din seria studiată nu au rezultat valori de clasare de 0 și 1 (analogii foarte apropiate), așa că am operat numai cu studiul valorilor analogice 2 și 3 (**Anexa 2**). Acestea se aranjează într-o tabelă de distribuție (tehnic, la origine, este un *query* de tip *crossstab*, în care pe rânduri găsim numerele ID ale „obiectelor-standard”, iar pe coloane „obiectele-țintă”). Această tabelă distribuțională a servit pentru o primă clasare a grupelor de compoziție. Grupele temporare au fost apoi aranjate într-o ordine cu sens (în ordinea descrescătoare a prezenței argintului în compoziție), rezultând grupele definitive, cele cu care se operează în **Anexa 3**.

Numele grupelor se compun dintr-o pereche de cifre (v. Anexa 1 sau 3), cu ordonanța descrescătoare a valorilor argintului, urmate de literele A sau B; aceste litere translează ideile anterior prezentate referitoare la analogiile cu valoare 2 și 3 (de mai sus), cu echivalențele 2=A și 3=B; la unele grupe, în special la cele numeroase, realizate printr-o încrângătură complicată de relații (de genul ID1 ~ ID2; ID2 ~ ID3; dar ID1 non~ID3), s-a simțit nevoia discriminării unor subgrupe, notate numeric, după A sau B. Rezultă clasificări de genul 01B1, 01B2,... 01B5, 02B1 etc, vezi **Anexa 3**).

Dintr-o perspectivă pur statistică rezultatele obținute sunt absolut credibile. De exemplu, chiar la grupa 1 (notată 01) avem tot tezaurul de la Teiuș, alături de o monedă ungurească ce a servit, foarte probabil, ca sursă de material, prin reciclare. Similar, cea mai mare parte a pieselor numismatice din Tezaurul de la Iași s-au grupat, majoritar, în grupa 5 (A și B), câteva piese fiind însă distribuite și în grupele 9B și 14A2, lucru absolut normal câtă vreme tezaurul nu sunt compuse doar dintr-o singură emisiune monetară (și cu atât mai puțin dintr-o singură șarjă metalurgică).

Comentariile pur numismatice am să le las în grija colegilor numismați, continuând aici cu chestiuni pur metodologice. Adjudicarea grupelor și subgrupelor tipologice, funcție de compoziția elementală (formula chimică exactă a fiecărui obiect), se lovește deocamdată de imponderabile, respectiv chestiuni procedurale care sunt încă de pus la punct. Problema se enunță, în cei mai simpli termeni cu putință, la stabilirea unor ecarteri de toleranță, care să discrimineze, cât mai exact, între noțiunile de *grupă* și *subgrupă*. Aceste ecarteri de toleranță sunt rezultatul unei serii de fapte prezentate în cele ce urmează.

Deși aparatele de măsură actuale sunt foarte precise, orice aparat în parte, dar și oricare procedură de măsurătoare sau evaluare, au erori tehnologice, erori admise între anumite limite. Asupra valorii acestor abateri (numite eufemistic *incertitudini*) există opinii divergente, dar situația se prezintă aproximativ în următorii termeni:

Există erori sistematice, pe care le generează chiar aparatul de măsură (fiindcă este decalibrat?), dar aceste erori ne interesează mai puțin aici, fiindcă derapajul este omogen (deci *sistematic*) și nu afectează judecata tipologică; mai mult, o eventuală recalibrare a aparatului poate duce la corecția în bloc a datelor obținute, fără eforturi substanțiale.

Există apoi erori nesistematice, care adică produc modificări aleatorii ale rezultatelor. Metoda folosită cu această ocazie (fluorescență de raze X) produce erori pe plaja de la 1,5% (în varianta optimistă) până la 5% (în varianta pesimistă). Evident, nici nu am să intru în alte detalii, nici nu am să deconspir sursele, lăsând specialiștii (fizicienii) să dezbată și să lămurească chestiunea. Personal, prin prisma rezultatelor obținute (care grupează corect tezaurul monetar, cel puțin aparent) tind să fiu... optimist și să dau credit, pentru moment, estimării erorii la 1,5%. Chiar dacă ar fi așa, înseamnă că, teoretic, două măsurători pe același obiect pot provoca diferențe de până la 3% (cu una la maximum și una la minimum). Iată deci de ce aplicația computerizată nici nu a sesizat analogii de nivel 1 (surprinse pe alte loturi testate). O a doua concluzie ar fi că analogiile de tip 2 pot fi considerate, prin corecție, ca fiind compoziții identice, deci posibil aceeași șarjă monetară. La limită, și unele analogii tip 3 ar putea fi în aceeași situație, dar, din nefericire, nu putem preciza care...

Un alt set de erori este generat de starea de conservare a pieselor, de procedurile concrete de restaurare și conservare, necesare, dar cu consecințe asupra măsurătorilor compoziționale; fișa tehnică de restaurare și conservare ar trebui să prevadă, *pentru fiecare obiect în parte*, procedurile urmate, substanțele folosite, pentru a permite o

evaluare cât mai puțin afectată de intervenții *moderne* asupra artefactului. Fiindcă asemenea date nu există, la nivelul solicitat, singura corecție posibilă este una experimentală, respectiv alegerea unui lot (loturi) de monede care să fie, cu mare probabilitate, aceeași emisiune monetară, pentru a măsura oscilațiile din interiorul ei, care ulterior să fie un etalon relativ de estimare a variabilității interne a unei emisiuni.

Un alt set de dificultăți în alocarea distribuțională este generat de chiar metrologia momentului istoric în cauză. Datorită inexistenței unor cântare foarte precise, cât și a imposibilității stabilirii compoziției reale a unor materiale reciclabile, șarjele metalurgice care furnizau material pentru una și aceeași emisiune monetară erau, fatalmente, non-identice. Soluția este, de asemenea, realizarea unor măsurători pe ambele fețe ale monedei; experimente realizate până acum au demonstrat că între avers și revers (ale aceluiași obiect) pot fi diferențe apreciabile, de până la 2%. O altă necesară corecție procedurală ar aceea că pe fișa pieselor trimise spre măsurare va trebui făcută mențiunea emisiunii unice (aceiași an de emisiune, aceeași matriță), tocmai pentru a putea stabili variabilitatea specifică a emisiunii.

Am să închei cu precizarea că comentariile de la Anexa 3 sunt făcute de mine, fără a poseda cunoștințe numismatice specifice și fără contact direct cu piesele. Am vrut doar să ilustrez modul în care datele furnizate de aplicația dedicată *metodei distanței medii* pot sugera concluzii de ordin istoric. Cu siguranță, acele comentarii vor trebui amendate de către numismați. Obiectivele mele au fost strict procedurale.

Anexele 1 și 3 au putut fi convertite la formatul Acrobat Reader (pdf), însă Anexa 2 a trebuit să o las în format Excel (unde datele pot fi vizualizate mai bine decât în Access, aplicația care furnizează datele de plecare), pentru că tabelul serial este foarte mare.