

**Departament de Prehistòria, Història Antiga i Arqueologia**

**Programa de Doctorat:**

**Ciències de l'Antiguitat (bienni 2003 – 2005)**

**Caracterització Arqueomètrica de la  
Ceràmica Vidrada Decorada de la Baixa  
Edat Mitjana al Renaixement als Centres  
Productors de la Península Ibèrica**

**Javier Garcia Iñáñez**

**Tesi Doctoral dirigida per:**

**Dr. Jaume Buxeda i Garrigós**

**Professor agregat d'Arqueologia**

# **7 Tècniques Estadístiques**



## **7.1 Els estudis de provenença**

Un dels problemes arqueològics més freqüents es troba a l'hora de diferenciar la provenença de la procedència de les ceràmiques o d'altres objectes trobats en un jaciment arqueològic. Primer cal, però, diferenciar dos conceptes que poden portar a la confusió, com són la procedència i la provenença. D'aquesta manera, la procedència es pot definir com el lloc de troballa de l'objecte subjecte d'estudi, mentre que la provenença és el lloc d'origen real de la matèria primera de la qual està fet l'objecte. Aquesta sovint coincideix amb l'origen de producció de l'objecte, tot i que no necessàriament. D'aquesta manera, i a mode d'exemple, una ceràmica trobada a Mataró procedeix d'aquest mateix lloc però, no obstant, podria provenir de Barcelona, és a dir, podria haver estat produïda a Barcelona i, posteriorment, transportada normalment a través de rutes comercials fins a Barcelona.

Els estudis de provenença es poden realitzar mitjançant dos mètodes: directe o indirecte. El mètode directe és aquell en que l'arqueòleg es basa en els seus coneixements per tal de, sense que se'n faci cap preparació específica, determinar i reconèixer les característiques de l'objecte que està estudiant. Aquest mètode es realitza mitjançant una observació macroscòpica i implica una vessant altament subjectiva en les seves interpretacions, esdevenint això una gran limitació (Buxeda *et al.* 1995, Buxeda 2001). Així un investigador pot suggerir, segons la seva experiència personal i l'observació de la ceràmica, que aquesta pertany a una producció determinada. No obstant, un altre investigador pot pensar que aquesta mateixa peça, en base a les seves atribucions estilístiques, per exemple, és en realitat una imitació de menor qualitat i per tant pertany a una altra producció. D'aquesta manera no es pot determinar qui és el investigador que té raó o, fins i tot, dir si ambdós es troben en un error d'atribució.

Només els estudis arqueomètrics, basats en l'aplicació de les ciències experimentals a l'arqueologia, poden superar aquesta limitació. Això és possible gràcies a que la provenença no ve determinada per un plantejament directe de la informació adquirida, sinó que l'investigador ha d'aplicar alguna tècnica analítica per tal d'obtenir la informació amb la que fer les interpretacions, tractant-se per tant d'un procediment indirecte. D'aquesta manera s'obté una informació que depèn només de la composició natural de la peça que s'analitza, així com del tractament d'aquestes dades, esdevenint una base objectiva i comparable per tal de proporcionar informació de la seva provenença, donant certa seguretat per a establir conclusions arqueològiques i històriques (Buxeda *et al.* 1995, Buxeda 2001).

## **7.2 La provenença a partir de l'anàlisi química.**

Actualment la majoria d'estudis de provenença es basen en la comparació de la caracterització química dels individus analitzats procedents dels centres receptors estudiats amb la d'una sèrie de Grups de Referència (GR) (Picon 1973) prèviament establerts formats a partir d'una sèrie significativa d'individus procedents de centres productors o fonts originàries d'extracció, coneguts i garantits. Aquests darrers GR permeten determinar les característiques pròpies de cadascun dels centres estudiats. En el present cas serien les característiques químiques de les ceràmiques majòliques produïdes en els principals centres productors de la Península Ibèrica.

En el cas de les ceràmiques, l'establiment dels GR presenten certes complicacions fonamentals. Aquestes es troben relacionades amb la matèria primera de què estan compostes les ceràmiques, és a dir, l'argila. Com ja van apuntar Widemann *et al.* (1975), tot i que referint-se a la *terra sigillata* de Lió però podent estendre's a d'altres produccions i èpoques, per a la majoria dels centres productors no es pot establir a on es troben les argileres naturals emprades pels ceramistes i ni tan sols saber si aquestes encara existeixen. A més a més, aquesta matèria primera ha patit sovint diversos processos de manipulació fins a la l'obtenció del producte final, podent així diferir substancialment la composició química de l'argila original de la ceràmica cuita. D'aquesta manera, l'atribució de la composició química obtinguda d'una ceràmica amb la de la seva font original d'argila pot esdevenir molt difícil. Per aquest motiu, la confecció d'un GR s'hauria de realitzar amb ceràmiques amb una atribució arqueològica el més segura possible. És per aquest motiu que s'ha difós ampliament el terme d' Unitats de Referència Composicional de Pasta (URCP) quan es tracta d'analitzar les pastes ceràmiques enlloc de matèries primeres, i en especial quan es tracta de ceràmiques que no procedeixen de tallers (Bishop *et al.* 1982).

La provenença mitjançant la caracterització química es basa en el conegut postulat de provenença:

*"...namely that there exist differences in chemical composition between different natural sources that exceed, in some recognisable way, the differences observed within a given source."* (Weigand *et al.* 1977).

El postulat de provenença estableix, doncs, que la variabilitat d'una mateixa font de matèria primera en principi ha d'ésser inferior a la variabilitat que podria existir entre

dues fonts diferents d'aquesta matèria primera. A més a més, es postula que aquestes diferències entre diverses fonts haurien de poder ésser identificables.

Els primers GR, anomenats als anys 60 grups de control, es feien sobre peces que arqueològicament es pensava que tenien una mateixa provinença:

*“A group for these purposes comprises similar sherds either found together or believed to have been made in the same locality” (Catling 1961).*

No obstant, sovint es prenen com a material segur ceràmiques trobades en un mateix jaciment arqueològic, assegurant la seva provinença sobre criteris subjectius, com són els criteris estilístics i morfològics. Per tant, l'error partia de base, ja que es podria estar analitzant ceràmiques que no necessàriament havien estat produïdes en el jaciment subjecte d'estudi. La manera d'evitar aquest tipus de problema en el cas de la caracterització de ceràmiques arqueològiques es soluciona mostrejant només peces d'atribució segura al centre de producció que volem caracteritzar. Això es tradueix, per exemple, en agafar per a anàlisis ceràmiques que s'hagin trobat a la zona de producció d'un taller ceràmic, especialment si aquestes presenten defectes de coccio. El fet que les ceràmiques trobades al taller, forns o abocadors de forns, presentin defectes de coccio significa que aquestes ceràmiques no deurien ésser comercialitzades, podent atribuir-les, en principi, un origen segur.

Un problema que pot esdevenir a causa de mostrejar ceràmiques amb defectes de coccio és que, potser, aquestes poden haver sofert canvis diferencials en la seva composició, podent no ésser característiques de les ceràmiques que sí es van comercialitzar durant l'activitat del taller que s'estudia. D'aquesta manera podríem caure en un greu error de caracterització, ja que les dades obtingudes de la mostra analitzada no podrien ésser inferides com a característiques reals de la població. Igualment, a l'hora de mostrejar ceràmiques amb defectes de coccio als voltants d'un centre de producció estem pressuposant que aquestes ceràmiques van ésser produïdes allà. No obstant, tot i que el més lògic és pensar això, es dona un cas peculiar quan aquestes ceràmiques amb defectes durant la coccio foren emprades per a la construcció. En aquests casos, les ceràmiques foren emprades sobretot a les voltes d'esglésies i edificis civils, presentant defectes produïts durant la coccio de les peces i seria lògic pensar que s'haurien produït a la ciutat a on es construeix l'obra. No obstant, es donen casos a on envasos de transport han estat amortitzats en construccions després del seu ús, com succeeix en construccions civils i religioses mexicanes i caribenyes amb

“botijas” espanyoles (Goggin 1962). Per tot això, cal afirmar que no sempre és oportú donar certes assumpcions com a certes, de manera que caldria un estudi per a cada cas concret.

Un nou model pels estudis de provinença basats en la caracterització arqueomètrica del materials, especialment ceràmiques, va ésser plantejat per Buxeda *et al.* (1995). Segons aquests autors, i tot seguint un diagrama de flux (Buxeda *et al.* 1995, Figura 4), un cop definit el problema arqueològic, es realitza el mostreig de les ceràmiques que es caracteritzaran arqueomètricament. Aquest mostreig ha d'ésser necessàriament aleatori, ja que s'ha d'evitar qualsevol discriminació subjectiva a l'hora de mostrejar els individus. A més a més, els autors proposen que aquest mostreig ha de partir d'una quantificació anterior basada en el nombre màxim d'individus (NMxI), ja que això únicament pot introduir un problema logístic en el treball arqueomètric en sobreestimar el número d'individus reals mostrejats, però mai el greu problema de subestimar el número real d'individus, el que portaria a un greu problema en la interpretació de les dades (per a una explicació més concisa sobre els mètodes de quantificació veure Buxeda *et al.* 1995, pp. 48-49).

Posteriorment, s'han d'identificar les diferents possibles fàbriques (F) existents en el conjunt d'individus ceràmics analitzats. En aquest cas, i tot seguint a Withbread (1989), entenem com a fàbrica la distribució, freqüència, forma, mida i composició dels diferents components d'una ceràmica. Un cop determinades les fàbriques, es poden identificar les pastes (P) que han donat com a resultat aquestes fàbriques. En aquest sentit, s'entén com a pasta la barreja d'argila o argiles, materials no plàstics, aigua, etc. que el ceramista empra per tal de produir les ceràmiques. El procés de preparació de la pasta correspon a un primer procés tecnològic, mentre que la fàbrica resulta d'un segon procés tecnològic, consistint bàsicament en el modelat i la cocció de les peces realitzades a partir de la pasta obtinguda en el primer (Buxeda *et al.* 1995). Aquest pas de pasta a fàbrica implica canvis químics (Kilikoglou *et al.* 1988), mineralògics (Maggetti 1981b) i microestructurals (Tite i Maniatis 1975, Buxeda *et al.* 1995).

No obstant, el procés real comença en la zona a on s'assentaria un taller (T), de la qual aquest captaria la matèria primera necessària. Aquesta zona és en realitat una zona d'incertesa ( $Z_i$ ), ja que al seu interior no és possible distingir analíticament entre les matèries primeres dels possibles diversos tallers emplaçats en ella, ja que aquestes presenten unes característiques geològiques iguals. A partir de la matèria primera d'una  $Z_i$ , un ceramista d'un T concret pot preparar diverses P, de la qual es podran obtenir

diferents F. Aquest mateix fenomen pot ocórrer repetidament en altres tallers situats a la mateixa  $Z_i$ , amb la qual cosa es donaran diverses P i, per tant, diferents F amb característiques geoquímiques semblants. D'aquesta manera s'afegeix una complexitat major a l'estudi arqueomètric, ja que pot fins i tot impossibilitar la diferenciació de diversos T procedents de la mateixa  $Z_i$  a partir de la caracterització dels seus individus ceràmics procedents dels centres productors, tot i poder-se establir  $Z_i$ . A més a més, caldrà tenir present les possibles contaminacions i/o alteracions que els individus ceràmics estudiats hagin pogut patir degut al seu ús i/o durant el seu període d'enterrament de manera no intencionada (Buxeda *et al.* 1995).

A més a més, els autors també fan ressaltar la possibilitat del que ells anomenen zones de conjunció ( $Z_c$ ). Aquestes presentarien característiques geològiques similars a  $Z_i$ , impossibilitant així la seva diferenciació analítica. D'aquesta manera pot ocórrer que dos T que produeixen el mateix tipus ceràmic presentin GR similars, tot i trobar-se ubicats en diferents  $Z_i$ . Aquest problema és més greu quan es treballa amb centres productors en lloc de tallers, amb els quals ens trobaríem en un nivell d'incertesa ( $N_i$ ). D'aquesta manera només es pot definir URCP, trobant-nos en un nivell de conjunció ( $N_c$ ) degut a la manca de coneixement directe sobre la o les  $Z_i$ . Només quan es pugui realitzar una associació de patrons entre una URCP i un GR ens trobarem en un  $N_i$  (Buxeda *et al.* 1995).

Com ja s'ha exposat, el gran ventall de possibilitats de que gaudeix un ceramista per tal de realitzar un tipus de ceràmica concreta, podent combinar diferents matèries primeres, i podent obtenir així diferents pastes, suposa que aquest pugui preparar en el mateix taller diversos tipus ceràmics amb funcionalitats diferents (ceràmica fina o de cuina, per exemple). D'aquesta manera, el resultat de l'estudi arqueomètric no serà la determinació d'un grup de control per a una àrea determinada, sinó la caracterització d'un grup de referència per a una producció determinada en una àrea concreta (Buxeda *et al.* 1995).

### **7.3 De les tècniques estadístiques univariants a les multivariants.**

Les tècniques estadístiques emprades en els moments inicials de l'arqueometria eren massa simplistes ja que es basaven en la utilització de tècniques descriptives univariants, és a dir, aquelles que consideren un conjunt de dades amb una sola



observació per cada un dels individus analitzats, per exemple els continguts de CaO. Els grups de control venien determinats per les mitjanes aritmètiques ( $\bar{x}$ ),

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

a on  $\sum x_i$  és el sumatori de la variable de cadascun dels individus que es volen tenir en compte a l'equació i  $n$  és el nombre d'individus emprats en aquesta, i desviacions estàndards ( $s$ ) de les concentracions de cada element químic determinat.

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}}$$

a on  $x_i$  és el valor de cada individu que compon la mostra,  $\bar{x}$  és la mitjana aritmètica i  $n$  el número d'individus que formen la mostra.

Aquests estudis assumien que les observacions de les dades composicionals seguien una funció de distribució normal (Buxeda 2001). Segons aquesta, el 68.46% dels individus del grup de control estarien entre -1 o +1 desviació estàndard respecte de la seva mitjana, el 95.46% entre -2 i +2 i, finalment, el 99.72% entre -3 i +3 desviacions estàndards. D'aquesta manera, els individus que quedaven enquadrats junts degut a la seva semblança química entre -3 i +3 desviacions estàndards respecte de la mitjana calculada del grup es determinava que formaven part d'aquest grup de control, mentre que la resta quedava fora. Tot i que els mètodes univariants poden ésser emprats per tal de calcular la tendència central d'una variable amb la seva corresponent dispersió, difícilment poden servir per a predir quina podria ésser la variabilitat que aquest grup de control pot presentar.

A partir dels anys 70, veient la necessitat d'interpretar les diverses variables dins d'un mateix grup de referència i no com a variables independents, es començà a aplicar les tècniques estadístiques bivariants, les quals prenen en compte dues variables alhora. Posteriorment s'aplicaren les anomenades tècniques estadístiques multivariants, les quals prenen en compte més de dues variables conjuntament.

La base de les tècniques bivariants i multivariants és que les diferents variables que es prenen en compte poden entendre's com a coordenades en un espai. Les

tècniques bivariants entenen un individu com un punt en un espai de dues dimensions, i la seva localització ve donada per les dues variables que es prenen en compte, essent els valors de dos elements químics en el cas de les anàlisis químiques de ceràmiques. Més enllà, les tècniques multivariants permeten superar les tres dimensions, treballant en l'espai  $n$ -dimensional, no essent possible la seva representació gràfica real. D'aquesta manera s'ha de recórrer a simplificacions de la realitat multidimensional mitjançant representacions en dues o, com a màxim, tres dimensions.

El concepte de distància és una de les pedres angulars en les tècniques multivariants. Si un individu es pot considerar com a vector, és a dir, com a un punt en un espai, determinat per les variables que el formen, considerant aquestes com a coordenades d'aquest espai, es podria mesurar quina és la distància que separa un individu d'un altre dins del mateix espai. D'aquesta manera, quant més gran sigui la distància entre aquests individus, menys semblants seran els dos individus entre ells mateixos. En el cas de la provenença emprant dades químiques, i tenint en compte com a coordenades els elements químics analitzats i els seus valors per a cada individu, quant més gran sigui la distància entre els dos individus, menys semblança química tindran, i a l'inrevés.

El càlcul de la distància entre dos individus a l'espai  $n$ -dimensional, no és gaire diferent del de la distància entre dos individus en un espai bidimensional. Tot i l'existència d'altres distàncies estadístiques, una de les més emprades actualment és la distància Euclidiana. Així, el càlcul de la distància Euclidiana és el càlcul de la hipotenusa a partir del teorema de Pitàgores, essent aquest l'arrel quadrada de la suma dels catets al quadrat. L'única diferència és que, enlloc de dos catets, en un espai multidimensional es disposa de  $n$  catets, tants com variables emprades com a coordenades al hiperespai. Per tant, si disposem d'una matriu de dades de  $m$  individus amb  $n$  variables cadascun que correspondran als elements químics analitzats, podrem determinar les distàncies que separen cada individu de la resta.

Les dades que utilitzem han de poder ésser treballades amb tècniques estadístiques inductives, en el sentit que treballem amb mostres d'individus molt més petites que la població real, inferint quines serien les característiques reals de la població. D'aquesta manera, s'assumeixen les funcions de densitat de la probabilitat, com ara la normal, per tal de poder utilitzar criteris probabilístics durant la investigació estadística. Com ja s'ha comentat, en arqueologia és gairebé impossible disposar d'un control total sobre les poblacions d'un subjecte d'estudi, esdevenint de vital importància

una bona inferència estadística de les dades mostrals per tal d'ajustar-se el més possible a les dades paramètriques reals.

En el tractament de les dades, s'han de tenir en compte altres coses, com és la variabilitat del grup composicional o de la població que s'estudia, que s'expressa a través de la variància,  $S^2$ , la qual correspon a la desviació estàndard al quadrat. La variància d'un grup, però, depèn de tres factors principalment. El primer és la variància natural ( $S^2_N$ ) que presenta la població que s'estudia, en aquest cas la variació natural de totes les majòliques que formarien un grup de referència. La segona és la variància mostral ( $S^2_S$ ), que va lligada al fet que no tots els materials arqueològics són homogenis. D'aquesta manera, si un espècimen no és gaire homogeni i agafem una mostra petita, pot donar-se el cas que els resultats obtinguts no siguin globalment representatius de l'individu. Això es pot evitar agafant mostres més grans per a cada individu i homogeneïtzant-les posteriorment per tal d'emprar una mostra representativa de la composició global de l'individu que s'analitza. Per últim, existeix la variància analítica ( $S^2_A$ ), que és aquella introduïda durant el processament dels individus mitjançant tècniques experimentals (errors de pesada, fluctuacions d'intensitats elèctriques que afecten les lectures analítiques dels aparells, precisió analítica, etc.). Per tal d'obtenir una variància total el més semblant a la variància natural s'ha d'intentar reduir al mínim la variància mostral i la variància analítica (Bieber *et al.* 1976, Glascock 1992). Així doncs, la variància total ( $S^2_T$ ) es pot expressar com:

$$S^2_T = S^2_N + S^2_S + S^2_A$$

L'aplicació de tècniques estadístiques en l'arqueologia suposa un cert perill de doble tall. Per una banda, es treballa amb mostres d'una població, ja que mai es podrà arribar a analitzar tots els individus que componen una població, com per exemple totes les ceràmiques d'un cert tipus que van existir en un període determinat, degut a que moltes no s'han descobert encara o potser mai es farà. D'aquesta manera, és impossible arribar a conèixer realment les atribucions de totes les majòliques produïdes, per exemple, al centre productor de Barcelona. D'altra banda, la ceràmica no és un objecte natural amb un tractament mínim és el cas de les obsidians per exemple, sinó que és un producte realitzat per l'home a partir d'una o més fonts naturals d'argiles i d'altres elements geològics i/o orgànics no plàstics, aigua, etc. D'aquesta manera, i com ja s'ha comentat, les característiques geoquímiques d'unes ceràmiques produïdes en el mateix

taller poden diferir sensiblement segons si s'han emprat diferents pastes degut a les necessitats o voluntats del ceramista. La gran majoria de ceràmiques es van realitzar mitjançant pastes preparades, esdevenint particulars aquells casos en que la matèria primera sense cap mena de preparació coincideix amb la pasta (Buxeda *et al.* 1995).

#### **7.4 El problema de les dades composicionals**

Les dades composicionals presenten un problema afegit que impedeixen aplicar els procediments estadístics habituals que s'empren per a la resta de dades. Aquest problema és que tots els valors es troben subjectes a la limitació de sumar 100 o la unitat. D'aquesta manera, el canvi d'una de les variables de la composició provoca també el canvi de la resta de variables. L'espai natural de les mostres de dades composicionals és el símplex de  $n$ -dimensions. En aquest sentit, s'ha de tenir en compte que la composició química d'una matèria primera pot canviar, ja sigui per la manipulació directa del ceramista durant el procés productiu (afegint desgreixadors, per exemple), o també com a efecte de diferents tecnologies productives amb la finalitat d'obtenir productes amb característiques diferenciades a través de la utilització de diferents temperatures de cocció o, també importants, a partir de coccions oxidants o reductores. En aquest sentit, durant la cocció d'una ceràmica s'eliminen, a mesura que augmenta la temperatura, l'aigua composicional (o química) de les argiles i altres components formats bàsicament per oxigen, hidrogen i carboni. Per tal d'avaluar la pèrdua d'aquests components durant la cocció original, es calcula en el laboratori allò conegut com amb el nom de Pèrdua al Foc (PAF), que consisteix en una calcinació controlada d'una part de la mostra. La PAF és directament proporcional a la temperatura de cocció, tot i que depèn de la composició inicial i de les fases secundàries. Per tant, quan més alta sigui la temperatura de cocció més alta serà la PAF, car una major quantitat d'elements volàtils s'hauran perdut. D'aquesta manera, la PAF influeix en l'augment o disminució de les concentracions dels elements determinats segons si aquesta és més alta o més baixa respectivament. Igual o més importants són els processos d'alteracions i/o contaminacions que poden patir les ceràmiques durant la seva etapa d'enterrament, tot i que encara sovint no reben el interès per part de la comunitat científica, donant-se els casos en que no es prenen en consideració en el tractament estadístic de les dades composicionals.

Una de les dificultats més aparents és la impossibilitat d'interpretar correctament les covariàncies i els coeficients de correlació de Pearson ( $r$ ):

$$r_{xy} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$$

Degut a la restricció que tots els valors tinguts en compte en la correlació sumen 100, es donen correlacions espúries que donen relacions de dependència falses, podent arribar així a interpretacions errònies. En principi, les covariàncies deurien poder adquirir lliurement valors nuls (0), positius o negatius, no obstant, degut a la restricció de sumar 100 aquestes covariàncies no són independents. Així:

$$\text{cov}(x_i, x_1) + \text{cov}(x_i, x_2) + \dots + \text{cov}(x_i, x_n) = 0$$

$$i = 1, 2, \dots, n$$

a causa de la restricció  $\sum_{i=1}^n x_i = k$  o *close to unit sum*. D'aquesta manera, sabem que  $\text{cov}(x_i, x_i) = \text{var}(x_i) > 0$ . Aquest fet provoca que hagi d'existir una covariància  $\text{cov}(x_i, x_j)$  ( $i \neq j$ ) de signe negatiu (Aitchison 1986, Mateu *et al.* 2003).

Un altre problema afegit a les dades, composicionals o no, és el pes que cada variable té en la composició. Per exemple, en una ceràmica analitzada, el SiO<sub>2</sub> o el Fe<sub>2</sub>O<sub>3</sub> presenten concentracions majoritàries, mentre que les concentracions de Th o Rb es troben a nivells de traces. D'aquesta manera, si calculem les distàncies directament sobre les concentracions determinades, el pes que tindran les concentracions dels elements majoritaris serà determinant en el càlcul de les seves distàncies. Per tal d'evitar tots aquests problemes s'utilitza el logaritme de les seves concentracions, quasi estandarditzant el pes de les diverses variables.

L'atenció, per tal de resoldre el problema de les dades composicionals, s'ha de centrar en els quocients de les variables  $x_i/x_j$  ( $i, j = 1, 2, \dots, n; i \neq j$ ). Treballant amb els quocients desapareixen els problemes de les correlacions espúries. A més, treballant amb els logaritmes naturals de les variables som subcomposicionalment coherents, ja que la magnitud relativa entre les parts d'una subcomposició no canvia en relació a la magnitud relativa entre les parts de la composició original, és a dir:  $s_i/s_j = x_i/x_j$  (Mateu *et al.* 2003). Aitchison basa la seva metodologia en la transformació de les dades composicionals a l'espai real multivariant, prenent així els logaritmes dels quocients, obtenint d'aquesta manera totes les dades en números reals amb els quals es pot aplicar

qualsevol tècnica estadística sense limitacions (Aitchison 1986, Aitchison *et al.* 2000, Mateu *et al.* 2003).

Existeixen diverses possibilitats de transformació de les dades basades en els logaritmes dels quocients entre les parts de dades composicionals, esdevenint les més usuals la transformació en logaritmes de quocients additius (alr):

$$\mathbf{y} = alr(\mathbf{x}) = \ln(x_1 / x_D), \ln(x_2 / x_D), \dots, \ln(x_{D-1} / x_D)$$

i la transformació en logaritmes de quocients centrades (clr):

$$\mathbf{z} = \ln(x_1 / g(\mathbf{x})), \ln(x_2 / g(\mathbf{x})), \dots, \ln(x_D / g(\mathbf{x}))$$

on  $D$  és el nombre de variables que es prenen en compte en el càlcul de cada individu i  $g$  és la mitjana geomètrica de tots els  $D$  components de  $\mathbf{X}$ , calculada segons:

$$g = \sqrt[D]{(x_1)(x_2)(x_3)\dots(x_n)}$$

Amb la transformació en logaritmes de quocients additives l'element utilitzat com a denominador adquireix un protagonisme especial respecte la resta, ja que la resta de variables seran dividides per ell. D'altra banda, la transformació en logaritmes centrada és simètrica entre les parts i la suma dels components transformats és igual a 0 (Aitchison 1986, Baxter 1989, 1991, 1992, Buxeda 1995, Buxeda *et al.* 1995, Aitchison *et al.* 2000, Mateu *et al.* 2003).

A les dades composicionals, l'estructura de covariància d'un grup  $\mathbf{X}$  de  $D$ -parts és el conjunt de totes les covariàncies:

$$\sigma_{ij,kl} = \text{cov}\{\log(x_i / x_k), \log(x_j / x_l)\} \quad (i, j, k, l = 1, \dots, D) \quad (\text{Aitchison 1986})$$

D'aquesta manera, les variàncies, les quals representen les variacions relatives entre dos components, resten definides per:

$$\tau_{i,j} = \sigma_{ii,jj} = \text{var}\{\log(x_i / x_j)\}$$

A més a més, l'estructura de covariàncies d'una composició  $x$  de  $D$ -parts està completament determinada per les  $1/2dD$  variàncies anteriors. D'aquesta manera, la matriu quadrada i simètrica  $D \times D$  és:

$$\mathbf{T} = [\tau_{i,j}] = [\text{var}\{\log(x_i / x_j)\}: i, j = 1, \dots, D]$$

la qual determina completament l'estructura de covariàncies, anomenant-se matriu de variació composicional (Aitchison 1986, Buxeda 1999b).

Un altre concepte important entorn de la matriu de variació composicional és el concepte de variació total ( $vt$ ), donat per:

$$vt = \frac{\sum_{i=1}^{i=D} \sum_{j=1}^{j=D} \tau_{i,j}}{2D}$$

Pràcticament, la variació total és la suma de totes les variàncies de la matriu de variació dividida entre per vegades  $D$ , que és el número d'elements determinats (Buxeda 1999b, Buxeda i Kilikoglou 2003). D'aquesta manera, la variació es troba relacionada amb les variàncies i covariàncies de l'estructura de covariàncies, quantificant la variabilitat continguda a les dades, en aquest cas químiques, i proporcionant així informació en part sobre la natura monogènica o poligènica d'aquestes.

Una altra distància emprada per molts investigadors en arqueometria és la Distància de Mahalanobis (DM) (Glascock 1992, Descantes *et al.* 2001, Descantes *et al.* 2004, Speakman *et al.* 2004). La DM mesura la diferència entre dues mitjanes (o centroides) de dos grups a l'espai multidimensional, així com també la d'un individu respecte d'un grup. Matemàticament, la DM d'un individu  $k$  respecte del centroide d'un grup  $A$  s'expressa:

$$D_{kA}^2 = \sum_{i=1}^n \sum_{j=1}^n [C_{ik} - A_i] I_{ij} [C_{jk} - A_j]$$

a on  $A_i$  i  $A_j$  són les concentracions mitjanes o elements  $i$  i  $j$  del grup i  $I_{ij}$  és l'enèsim element de la inversa de la matriu de variàncies-covariàncies. La DM permet càlculs de

la probabilitat que un individu particular pertanyi a un grup basat no només en la seva proximitat al centroide del grup en termes de distància euclidiana, sinó també segons el grau en que la densitat dels punts dels individus disminueixen des del centroide del grup cap l'individu que s'avalua. La significància de les diferències entre dos grups d'individus pot ésser comprovada mitjançant la  $T^2$  de Hotelling (l'equivalent multivariant de la  $t$  d'Student) segons:

$$T^2 = \frac{D^2}{\frac{1}{n_1} + \frac{1}{n_2}}$$

a on  $n_1$  i  $n_2$  són el nombre d'individus en els dos grups. La probabilitat de pertinença és fàcilment calculada a partir de la  $T^2$  emprant el valor  $F$  segons:

$$F = \frac{[n_1 + n_2 - v - 1]}{[n_1 + n_2 - 2]v} T^2$$

a on  $v$  és el nombre d'elements de l'anàlisi. No obstant, el càlcul de la DM necessita que el número d'individus de cada grup tingut en consideració sigui, al menys, una unitat superior al número d'elements emprats. En condicions ideals, el número de membres de cada grup serà diverses vegades més gran que el número d'elements (Glascock 1992).

Per tal de representar gràficament les distàncies químiques obtingudes entre els individus s'utilitzen les esmentades tècniques de reducció de la dimensió, esdevenint l'anàlisi de conglomerats o *cluster analysis*, una de les més emprades en els estudis de provinença. Cal afegir que en el present treball s'utilitza l'anàlisi de conglomerats mitjançant el mètode aglomeratiu del centroide. En aquesta tècnica, l'origen de la qual s'ha de buscar en la taxonomia numèrica biològica, el gràfic resultant es realitza en forma de dendrograma o gràfic d'arbre. A partir de la matriu de distàncies obtinguda de les dades químiques, es situen tots els individus analitzats a la base del dendrograma i es van unint entre ells fins que resten tots connectats. La distància a la que dos individus s'ajunten ve donada pel punt més alt respecte de la base del dendrograma, de les línies que els connecten. Aquesta línia, doncs, es pot mesurar amb l'escala de distàncies



ultramètriques que es troba a l'esquerra del dendrograma. Contra major és aquesta distància, menor és la similitud química (Sokal i Sneath 1973, Buxeda 1999b, 2001).

L'existència de possibles grups d'individus en l'espai descrit per la matriu de dades és el que s'anomena l'estructura de la matriu de dades. Aquesta és l'ocurrència diferencial de punts de dades en l'espai n-dimensional definit per les concentracions elementals (Bishop i Neff 1989). Degut a que les diverses tècniques estadístiques bivariants i multivariants estan centrades en la recerca de descripció de les variables, les seves relacions i les possibles estructures que presentin les dades contingudes en la matriu de distàncies, podem dir que són només mètodes descriptius, no pas explicatius, ja que el resultat del tractament estadístic és un resum de les dades per tal de permetre un estudi organitzat de les mateixes.

Altres tècniques s'empren comunament són l'Anàlisi de Components Principals (ACP) i l'Anàlisi Discriminant (AD). L'ACP és una tècnica estadística de síntesi de la informació o de reducció de la dimensió (en aquest cas del número de variables). Això significa que, davant d'una sèrie de dades amb moltes variables, l'objectiu de l'ACP serà reduir-les a un nombre inferior perdent la menor quantitat possible d'informació. Els nous components principals seran així una combinació lineal de les variables originals i, a més, seran independents entre si. Això ens permetrà observar la relació existent entre els grups amb els diversos components i, si es situen en el mateix gràfic els components originals, en aquest cas els elements químics, es podrà observar les principals diferències o similituds entre els individus analitzats. En altres paraules, l'ACP determina la direcció i magnitud de la màxima variació d'un conjunt de dades a l'espai n-dimensional. Per aquest motiu, l'ACP és extremadament útil quan s'utilitza sobre dades composicionals amb una alta correlació (Davis 1986). En canvi, l'AD suposa que es poden dividir les dades en diferents grups segons uns criteris previs, i a continuació intentar trobar una forma de distingir entre aquests grups calculant la contribució de cada una de les variables originals en la diferenciació dels grups establerts (Davis 1986, Shennan 1992).

### **7.5 Altres aproximacions al tractament de les dades composicionals**

El tractament estadístic de les dades composicionals no és homogeni per part dels investigadors que treballen en els diversos camps de l'arqueometria. D'aquesta manera s'apliquen diferents models estadístics per tal d'estudiar les dades i revelar les

seves possibles estructures. És per això que Baxter (2001) ha realitzat un estudi comparatiu entre tres tendències d'apropament a les dades composicionals, Brookhaven/MURR, Bonn i Barcelona, les quals, tot i semblar diferents, arriben a oferir resultats similars. El procediment estadístic emprat al laboratori de Brookhaven i, per extensió, al MURR, resulta en la transformació de les dades composicionals en logaritmes en base 10 abans de qualsevol altre operació (Bieber *et al.* 1976, Lizee *et al.* 1995, Steponaitis *et al.* 1996, Hegmon *et al.* 1997, Triadan *et al.* 1997, Herrera *et al.* 1999). Això és degut a que s'assumeix que les variables (sovint elements traça en ppm) tenen una probabilitat més alta que es trobin en una distribució normal multivariant dins dels grups de l'escala logarítmica, essent a més aquesta distribució normal multivariant necessària per als procediments emprats amb posterioritat. Sovint es considera que l'error analític no és important en relació a la variació natural de la mostra, essent aquest finalment ignorat, com ho justifica Bieber *et al.* (1976). Una de les principals preocupacions és el problema que exhibeixen les correlacions entre les variables d'una població. Si aquestes apareixen d'una manera alta, donaran lloc a poblacions que tindran una forma el·líptica a l'espai n-dimensional. D'aquesta manera, l'anàlisi d'agrupament no es considera adient, ja que tendeix a produir agrupament esfèrics de mida similar que poden desorientar si l'estructura real de la mostra és el·lipsoïdal. Per aquest motiu, Brookhaven/MURR releguen l'anàlisi d'agrupament a un segon pla, aplicant una metodologia d'avaluació de grups per tal de definir provisionalment els possibles grups o estructures. Posteriorment, i assumint que els grups establerts preliminarment estan constituïts per mostres que segueixen una distribució normal multivariant, es calcula la probabilitat de pertinença a un grup a partir de la distància de Mahalanobis, que permet que es formin agrupacions el·lipsoïdals. Quan aquest procediment no és immediatament possible, Brookhaven/MURR utilitzen la tècnica de reducció de variables de l'anàlisi de Components Principals o, també, un subgrup de variables que semblen discriminar entre els grups. Els individus són afegits o extrets d'un grup fins que s'aconsegueix un agrupament estable. Cal remarcar que l'assumpció de complir la distribució normal multivariant implica rebutjar tots aquells casos en que aquesta no és compleix, tendint així a imposar una estructura d'aquest tipus. El grup de Bonn (Mommsen *et al.* 1988, Beier i Mommsen 1994), no obstant, qüestiona la necessitat d'emprar dades transformades en logaritmes, la manca d'importància de l'error analític, així com la importància que altres investigadors donen a les altes correlacions entre variables. La idea central del seu procediment, a més de l'anomenat

“*best relative fit*” consisten en un càlcul de la seva “dilució” i la seva correcció, radica en fer créixer un grup a partir d’un únic cas o d’un nombre petit de casos similars, als quals se’ls hi anirà afegint aquells casos que siguin més semblants. Aquest procés continua invariablement fins que cap altre individu pot afegir-se al grup originari, moment en el qual es crea un nou grup des d’un altre punt repetint-se el procés, i així fins que tots els individus similars s’ajuntin, definint-se com a *outliers* aquelles que no quedin enquadrades en cap grup definit. La pertinença a un grup determinat es decideix a partir dels càlculs de distància i de probabilitat, recolzat per l’assumpció d’una distribució normal multivariant, i emprant tan la distància Euclidiana com la distància de Mahalanobis. El fet de no emprar dades transformades logarítmicament és defensat pel grup de Bonn argüint que els resultats del treball estadístic amb les dades no transformades és similar. No obstant, tal com Baxter (2001) apunta, aquest fet es pot donar en el cas de ceràmiques molt fines i estandarditzades, no podent-se generalitzar. Per últim, el tractament de les dades emprat a Barcelona (Buxeda 1999b, 1999a, 2001, Buxeda i Kilikoglou 2003), seguint els treballs d’Aitchison (Aitchison 1983, 1986, Aitchison *et al.* 2000), proposa emprar les dades transformades en logaritmes de raons a partir d’una transformació asimètrica, en la qual una de les variables serà escollida com a divisor per a la resta. L’elecció del divisor es realitza a partir de l’observació del comportament de cada variable quan són utilitzats com a divisors en la matriu de variació composicional. Així, la variable que imposi una variabilitat menor a la variació total de la matriu serà escollit com a divisor. Un cop les dades han estat transformades en logaritmes de raons, aquestes es treballaran a partir de l’aplicació de tècniques estadístiques tan multivariants com bivariants o univariants. Cal afegir que és aquesta darrera aproximació estadística l’emprada en aquest treball.