

Departament de Prehistòria, Història Antiga i Arqueologia
Universitat de Barcelona

Programa de Doctorat:
Ciències de l'Antiguitat (bienni 2003-2005)

**Estudi arqueomètric i arqueològic de la
producció i difusió d'àmfores vinàries de la
zona central i sud de la costa catalana
durant els segles I aC i I dC**

Verònica Martínez Ferreras

Tesi Doctoral dirigida per:

Dr. Jaume Buxeda i Garrigós
Professor agregat d'Arqueologia

Capítol 5

TRACTAMENT ESTADÍSTIC DE LES DADES COMPOSICIONALS

Paral·lelament al desenvolupament i estandardització dels mètodes fisico-químics d'anàlisi, s'ha avançat en els principis i les tècniques estadístiques per al tractament del gran volum de dades que generen les anàlisis arqueomètriques.

En els estudis de provenença, la finalitat del tractament estadístic de les dades composicionals és la seva exploració, ordenació i esquematització, per tal d'interpretar les relacions composicionals entre els individus ceràmics en estudi. El mètode estadístic escollit i els elements representats en el tractament estadístic dependrà de les assumpcions formulades amb anterioritat per tal d'identificar l'estructura existent en el conjunt de dades (Bishop i Neff, 1989: 57). Donada la importància de la variabilitat en la determinació de la provenença mitjançant l'anàlisi química, que ve expressada en el Postulat de Provenença, és important, en l'anàlisi estadística, avaluar els graus de variabilitat existents en el conjunt de dades en la definició dels grups de referència (GR).

La caracterització química permet determinar la concentració dels elements majors i menors, que venen expressats en percentatges d'òxids, i dels elements traça, expressats en parts per milió (ppm) (1ppm equival a 0.0001%). Les diferents concentracions químiques determinades, Fe_2O_3 , Al_2O_3 , MnO , P_2O_5 , TiO_2 , MgO , CaO , Na_2O , K_2O , SiO_2 , Ba, Rb, Th, Nb, Pb, Zr, Y, Sr, Sn, Ce, Co, Ga, V, Zn, Cu, Ni i Cr, seran les variables definides per a cada un dels individus ceràmics en estudi. Aquests valors venen representats en forma d'una taula, una matriu de dades $n \times D$, en la que el conjunt de files representen els individus en estudi, representats amb una i , i cada columna representa les variables, les concentracions elementals, representades amb una j . A les cel·les trobem els valors de cada individu per a cada variable, expressats amb una x , que representaran els valors de 1 a n individus, i de 1 a D variables, ja que n correspon al nombre total d'individus i D al nombre total de variables. Les composicions elementals determinades són observacions numèriques d'uns individus analitzats. Depenent de la quantitat de variables o atributs d'interès del nostre estudi, les tècniques estadístiques utilitzades seran univariants (1 variable), bivariants (2 variables) o multivariant (més de 2 variables).

En el cas dels estudis de provenença sobre ceràmica arqueològica, com que treballarem només amb una part dels individus d'una població, els valors composicionals d'un GR seran considerats com estimacions dels paràmetres que venen definits, en primer lloc, per la mitjana aritmètica (μ) de cada variable, com a mesura de tendència

central, que ens indica al voltant de quin valor mitjà es troba la concentració en els diferents individus i la desviació estàndard (Shennan, 1988; Baxter, 1994):

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, \text{ o } \bar{x} = \left(\frac{x_1}{n}\right) + \left(\frac{x_2}{n}\right) + \dots + \left(\frac{x_n}{n}\right),$$

on $\sum x_i$ correspon al sumatori de la variable de tots els individus i n equival al nombre d'individus. A més, es calcula la variància (s^2), és a dir, quant es desvia cada concentració del valor de la mitjana però, com que la suma d'aquestes desviacions dóna sempre 0, el valor obtingut s'eleva al quadrat, de manera que tots els valors siguin positius i finalment es divideix pel nombre d'individus. La seva arrel quadrada és la desviació estàndard (s), com a mesura de dispersió al voltant de la mitjana aritmètica (Baxter, 1994), segons

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

on x_i correspon al valor de cada individu.

Les tècniques bivariants es basen en la situació que presenten els individus en relació a un plànol en el qual les dues coordenades venen representades per les dues variables considerades (Figura 5.1). Els tres individus, A, B i C, es situen com a punts en aquest espai bidimensional, segons els valors que representen per a cada coordenada. Aquesta representació geomètrica permet observar les relacions entre els individus segons la seva posició. Si unim els punts amb una línia, aquesta reflectirà l'existència d'una determinada relació entre ells. D'aquesta manera, en la figura 5.1, la distància entre l'individu A i l'individu B pot ser calculada, ja que la unió de les seves coordenades dibuixa un triangle rectangle. Si, per exemple, l'eix d'abscisses ve representat pels valors de MgO i l'eix d'ordenades pels valors de CaO dels 3 individus considerats, A, B i C, el catet paral·lel a l'eix d'abscisses marcarà les diferències de MgO en els individus, mentre que el catet paral·lel a l'eix d'ordenades representarà la diferència dels valors de CaO. La distància (d) que els separa, corresponent a les seves diferències composicionals en MgO i CaO, ve representada per la hipotenusa, que es pot

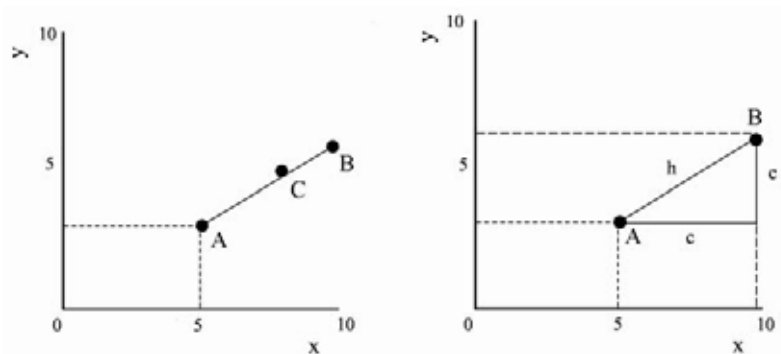


Figura 5.1: Gràfic bivariant representant la distància que separa als individus A i B

calcular mitjançant el teorema de Pitàgores, $h^2 = c^2 + c^2$, i que correspondrà a la distància euclidiana. Si treballem amb un nombre major d'individus, les distàncies entre tots ells vindran marcades per les diferències en els costats.

Per tal de determinar la relació entre dues variables s'utilitzen les covariàncies (Baxter, 1994):

$$\text{cov}_{i,j} = \frac{1}{(n-1)} \sum_{k=1}^n (x_{ki} - \bar{x}_i)(x_{kj} - \bar{x}_j),$$

on x_{ki} és el valor de la variable i en l'individu k , mentre que x_{kj} és el valor de la variable j en l'individu k .

Les covariàncies poden ser representades en una matriu de covariàncies o de variància-covariància, que serà simètrica i quadrada. El grau de la relació entre dues variables, la correlació, es mesura amb el coeficient de correlació lineal producte moment de Pearson (r), que es mesura a partir de les covariàncies, proporcionant una matriu de correlacions. A partir de les variables x i y la representació del coeficient de correlació seria

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

Els valors del coeficient de Pearson van de -1 (correlació inversa perfecta) a +1 (correlació directa perfecta), essent 0 la independència o no correlació.

La definició dels grups de referència implica el control de totes les fonts de variabilitat que poden afectar les concentracions químiques dels individus ceràmics (Buxeda, 1999a, 1999b; 2001; Buxeda *et al.*, 1995; Buxeda i Kilikoglou; 2003). Tanmateix, aquestes estimacions són limitades ja que la gran variabilitat composicional existent entre els individus pot deure's a diferents factors. Fins als anys 1980, la variància ha estat definida com la variància total dels individus ceràmics (S_T^2), que és la suma de la variància natural (S_N^2) de l'àrea de la font geològica natural, de la variància del mostreig (S_S^2), deguda a la homogeneïtat de la mostra i de la variància analítica (S_A^2), que depèn de la precisió i de les condicions dels mètodes analítics (Bieber *et al.*, 1976; Weigand *et al.*, 1977). A partir dels anys 1980, s'incorporaren altres factors que poden introduir una certa variabilitat al conjunt composicional, com poden ser les estratègies d'aprovisionament dels diferents ceramistes, la diversitat de processos tecnològics que aquests poden haver utilitzat en la cadena operacional de la manufactura i les condicions d'ús i de deposició de les ceràmiques (Buxeda *et al.*, 1995).

Per tal de superar les dues i tres dimensions a les que està restringit el nostre espai geomètric, el tractament estadístic de les dades composicionals requereix l'ús de tècniques estadístiques multivariants. En superar les tres dimensions podem situar les dades composicionals dels individus ceràmics a l'hiperespai o espai n-dimensional, en el que D equival al nombre de dimensions o variables que es fan servir en cada cas.

La distància euclidiana, però, no pot ser calculada directament sobre les dades composicionals ja que existeixen diferències en la concentració dels elements majors i menors, que tenen un pes important, i traça, que es veurien ínfimament representats. L'equilibri en quant al pes de les diverses variables es pot aconseguir, per exemple, amb l'estandardització de les concentracions elementals

$$z_i = \frac{(x_i - \bar{x})}{s}$$

La concentració determinada es substitueix per la puntuació tipificada, que és la desviació de la concentració inicial respecte a la mitjana de les concentracions que presenta aquest element entre els individus, dividit per la seva desviació estàndard.

D'aquesta manera s'aconsegueix que les variables estandarditzades tinguin totes una mitjana de 0 i una desviació estàndard d'1.

En un espai n -dimensional, també es pot calcular el valor de la hipotenusa, és a dir, la distància euclidiana (Baxter, 1994). Hi haurà tants catets com D variables emprades i tots es calculen com les diferències en les concentracions que presenten els individus per a cada element (n). Permet calcular la distància que separa a tots aquests individus i veure quins estan més o menys a prop en aquest espai n -dimensional. El resultat del tractament estadístic és una matriu de distàncies quadrada, ja que té el mateix nombre de files que de columnes. Les cel·les representen la distància que l'individu d'aquella columna té respecte l'individu d'aquella fila. A més, la diagonal central, formada per la distància entre la fila 1 i la columna 1, entre la fila 2 i la columna 2, etc., tenen el valor de 0, ja que no hi ha cap distància entre un individu i ell mateix. I, finalment, el triangle de distàncies que es troba per sota de la diagonal és simètric del triangle que es troba per sobre de la diagonal.

Un problema que presenten les dades composicionals és que expressen una part del total i aquest total, s'expressi com s'expressi (% , freqüències relatives, etc.), està subjecte a una constant, k , que en el cas de les dades composicionals d'un individu ceràmic podria ser 100%. Aquesta restricció s'anomena "restricció de suma a la unitat" i crea una dependència entre les variables, el valor de les quals correspondrà a la diferència entre 100 i la suma del total de variables considerades (Aitchison, 1986; Buxeda, 1999a, 2001). Tot i que en el cas de l'anàlisi de dades composicionals la constant és el 100% de la suma de les variables, aquest deuria representar a totes les composicions elementals presents en l'individu però, en canvi, només es tenen en consideració certs elements. Per aquest motiu, el resultat de l'anàlisi química afecta només a una subcomposició, que acostuma a venir representada per uns 25-30 elements químics. Un altre problema ve expressat per l'efecte de dilució que pot afectar a un grup d'individus que han estat fabricats amb la mateixa argila però amb diferent tipus de desgreixador. A més, altres processos tecnològics de fabricació poden contribuir a alterar la composició química de dos individus fabricats amb la mateixa matèria primera original. Durant la cocció, per exemple, es perden certs components que, depenent del grau de temperatura assolida, l'atmosfera i el temps de cocció, implicarà diferents graus de pèrdua i d'alteració. Per evitar aquest problema, una de les vies és la normalització de les dades, és a dir, calcular les concentracions determinades per a que la concentració de tots els individus sigui 100%. Aquest càlcul produirà nous valors que són

independents de les composicions no determinades i dels components del vector de pertorbació no determinats però, en canvi, el fet de multiplicar cada component de la subcomposició determinada per un factor d'escala, comporta certs límits (Buxeda, 1999a). La dependència a una suma constant fa que l'espai en el que es situen les dades composicionals no sigui realment un espai euclidià real de D -dimensions. Igualment, aquesta restricció fa que no es pugui interpretar adequadament la intensitat de la relació entre les variàncies, és a dir, les covariàncies i els coeficients de correlació de Pearson, ja que són restringides i falsegen els resultats. Per aquest motiu, s'anomenen correlacions espúries.

Per tal de superar la restricció a la constant de les dades composicionals, Aitchison (1982, 1985, 1986) va elaborar un cos teòric associat a les propietats algebraiques i geomètriques de l'espai mostral, on la suma de les D -parts considerades ve representada pel símplex S^D , i que depèn de la constant (k):

$$S^D = \left\{ (x_1, x_2, \dots, x_D)' : x_i > 0; \sum_{i=1}^D x_i = k \right\}.$$

Aquest desenvolupament teòric es fonamenta en la idea que les dades composicionals són aquelles que expressen part d'un total i els resultats responen als valors relatius entre les composicions. D'aquesta manera, les dades composicionals es transformen en concentracions relatives (divisions o raons), ja que qualsevol composició \mathbf{x} ($\mathbf{x} = x_1, \dots, x_D$) ve determinada per les d raons dels components x_i/x_D ($i=1, \dots, d$; $d=D-1$). Finalment, es va adoptar la proposta de treballar amb el logaritme d'aquestes raons, mitjançant la transformació en logaritmes de raons. Així, es produeix una quasi estandardització de les variables transformades i totes les variables presenten un pes similar en el càlcul de la distància (Aitchison 1982, 1983, 1986, 1992; Aitchison *et al.*, 2002; Buxeda, 1999a; Mateu *et al.*, 2003). Mitjançant la transformació de les dades composicionals en logaritmes de raons centrats, obtenim uns nous valors transformats en concentracions relatives que ja no estan subjectes a cap restricció i, a més, no variaran en relació als valors relatius de la composició original. La transformació de logaritmes de raons centrats asimètrica es basa en:

$$\mathbf{x} \in S^d \rightarrow \mathbf{y} = \ln\left(\frac{\mathbf{x}_{-D}}{x_D}\right) \in \mathbb{R}^d,$$

on $\mathbf{x}_D = (x_1, \dots, x_D)$, utilitza una de les composicions com a divisor en la transformació de logaritmes i la suma d'aquests logaritmes dependrà d'aquest divisor. D'aquesta manera, podem estimar quins elements són susceptibles d'estar pertorbats (Buxeda, 1999a: 299). Per tal de representar les dades composicionals en un espai multivariant, es pot utilitzar la distància euclidiana al quadrat, ja que els valors relatius no estan subjectes a cap restricció.

Per tal de definir la variabilitat dels valors composicionals relatius, Aitchison va definir la subcomposició basada en parts (1, 2, ..., C) d'una composició de D -parts (x_1, \dots, x_D) com la subcomposició-(1, 2, ..., C), definida per:

$$(s_1, \dots, s_C) = (x_1, \dots, x_C) / (x_1 + \dots + x_C).$$

Aitchison introdueix la relació de les parts d'una composició i l'estructura de la covariància d'una composició \mathbf{x} composta de D -parts, com el grup de totes les covariàncies, que podem expressar en una taula $D^2 \times D^2$, segons (Aitchison, 1986: 65; Buxeda, 1999a: 301):

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

on cada variància ve definida pels $\frac{1}{2}dD$ valors $\tau_{ij} (i = 1, \dots, d; j = i + 1, \dots, D)$, com:

$$\tau_{ij} = \sigma_{ii,jj} = \text{var}\{\log(x_i / x_j)\} \quad (i, j = 1, \dots, D).$$

En canvi, l'estructura de la covariància d'una composició \mathbf{x} de D -parts es determina mitjançant les variàncies prèvies $\frac{1}{2}dD$

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

La matriu que exemplifica aquestes variàncies determina completament l'estructura de la covariància i s'anomena matriu de variació composicional (MVC) (Aitchison, 1986, 1992; Buxeda, 1999a; Buxeda i Kilikoglou, 2003). Aquesta matriu és simètrica i quadrada $D \times D$. Paral·lelament, Buxeda (1999a: 301) contribueix a aquest

desenvolupament teòric indicant que, en aplicar la transformació de logaritmes en una subcomposició de S-parts, es produeix la relació $\tau_{is} = \sigma_{ii,ss}$, de tal manera que els valors de la columna S-èsima de la MVC són iguals als de la diagonal de la matriu de covariàncies resultant d'utilitzar un element químic (x_s) com a divisor

$$\Sigma = [\sigma_{ij,ss}] = [\text{cov}\{\log(x_i/x_s), \log(x_j/x_s)\}] \quad (i, j = 1, \dots, S).$$

Però aquesta correspondència només és parcial, perquè la matriu de covariància logarítmica no conté la variància τ_{ss} , perquè disposa d'una dimensió menys (la variable s utilitzada com a divisor) i, com que $\tau_{ss} = 0$, podem calcular la variància total d'una matriu de covariància logarítmica, com:

$$\tau_{\bullet s} = \sum_{i=1}^S \tau_{is}.$$

Així doncs, la MVC ens indica la variabilitat introduïda per cada element químic al conjunt de dades ($\tau_{\bullet i}$). A més calcula la variació total (vt) del conjunt de composicions per al total d'individus (n) (Aitchison 1986; 1990; Buxeda, 1999a; Buxeda i Kilikoglou, 2003):

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

que també podem formular: $vt = (2S)^{-1} \mathbf{j}' \mathbf{T} \mathbf{j}$, on \mathbf{j} és el vector $S \times 1$ d'unitats. La vt és, doncs, la suma de totes les variàncies en la matriu de variació dividides per dues vegades S , el nombre d'elements determinats. Així, la variació total està relacionada amb les variàncies i covariàncies de l'estructura de covariància i quantifica la variabilitat continguda en un conjunt de dades. A més, la vt de les dades logarítmiques transformades està directament relacionada amb la distància euclidiana entre tots els espècimens (Aitchison, 1992; Buxeda i Kilikoglou, 2003). D'aquesta manera, la variació total dóna una mesura de variabilitat en el conjunt de dades químiques, en conformitat tant amb la variància (traça) com amb la distància Euclidiana. Amb aquest càlcul podem estimar la $\tau_{\bullet i}$, que és la suma del total de variàncies transformades de cada

columna (i) i la freqüència relativa que suposa la v_i respecte el sumatori de les variàncies de cada columna ($v_i/\tau_{.i}$).

Per tal d'evitar els problemes de pertorbació dels individus ceràmics com a conseqüència de possibles efectes d'alteració i/o contaminació durant els processos de fabricació, ús, deposició i post-deposició, Buxeda ha elaborat un model alternatiu multiplicatiu en el que la pertorbació pot ser definida com (Aitchison, 1986; Buxeda, 1999a):

$$\mathbf{x}_1 = \mathbf{u}_1 \circ \mathbf{x}_0 = C(x_{01}, u_{11}, \dots, x_{0D}, u_{1D}).$$

En aquesta operació, les D-parts del vector composicional \mathbf{x}_0 , poden ser identificades mitjançant el vector de la composició original de la ceràmica, que és operat per les D-parts del vector pertorbat \mathbf{u}_1 d'elements positius. D'aquesta manera, obtenim el vector composicional pertorbat \mathbf{x}_1 . C representa l'operador de closura, que transforma cada vector de D-parts en un vector que representa la suma de la unitat. En el cas de treballar amb una seqüència de composicions generades per pertorbacions successives, aplicaríem (Aitchison, 1992; Buxeda, 1999a):

$$\mathbf{x}_n = \mathbf{u}_n \circ \mathbf{x} [\mathbf{x} = \mathbf{x}_0; \mathbf{u} = (\mathbf{u}_1 \circ \dots \circ \mathbf{u}_n)], (n=1, 2, \dots).$$

D'aquesta manera, es comprova que l'operació de pertorbació en dades composicionals es pot establir o bé directament, on el component x_j és multiplicat pel component u_j , o indirectament, on el component x_j és dividit per k , que és una pertorbació imposada per l'operador de closura. Així, les possibles fonts de variació en D-parts de \mathbf{x}_0 són tant les pertorbacions directes com indirectes dels operadors de pertorbació (Buxeda, 1999a).

Baxter (2001: 139) suggereix que, tot i que el model proposat per Buxeda és heurístic, l'evidència empírica el fa vàlid, però considera que es poden obtenir els mateixos resultats treballant amb dades logarítmiques transformades no estandarditzades.

Existeixen diferents mètodes de representar gràficament el càlcul de les distàncies de dades composicionals, com l'anàlisi de conglomerats (*cluster analysis*), l'anàlisi de components principals o l'anàlisi discriminant, entre d'altres (Baxter, 2006).

Els diferents mètodes d'anàlisi de conglomerats es basen en models heurístics que segueixen una funció de densitat normal en la seva versió multivariant. Impliquen, d'una banda, la selecció d'una mesura de (dis)similaritat entre parells d'individus d'una

població que ha de ser agrupada i, d'altra banda, la selecció d'un algoritme per agrupar els individus jeràrquicament, partint de la base del coeficient de (dis)similaritat. Tanmateix, els grups i els membres de cada grup es basen en estimacions de paràmetres desconeguts, les assumpcions dels quals prenen la forma d'una funció de densitat probabilística, per això l'aproximació és heurística. El model derivat d'aquestes assumpcions ens permet l'ús d'inferències estadístiques i de càlculs probabilístics. D'aquesta manera, si l'anàlisi consisteix en l'observació d'un vector (una població d'individus) de D variables, la distribució normal multivariant vindrà definida per un vector de mitjanes aritmètiques desconegut i una matriu de covariàncies. El nombre de grups (G) és inicialment desconegut i la finalitat serà estimar la probabilitat que un individu pertanyi a un grup, avaluar si un grup és químicament coherent, així com determinar el nombre de grups i els individus que conformen cada grup. Tanmateix, es considera que els mètodes d'agrupament heurístics imposen una estructura esfèrica de les dades que pot impossibilitar la identificació de l'estructura real (Papageorgiou *et al.* 2001).

L'anàlisi de conglomerats, que és descriptiva, no explicativa, utilitzada en el nostre estudi és el mètode aglomeratiu del centroide. La representació d'aquest mètode és el dendrograma, la forma del qual depèn de l'escala de les dades, la mesura de (dis)similaritat escollida (per exemple, la distància euclidiana, la distància de Mahalanobis, etc.) i l'algoritme d'agrupament utilitzat (per exemple, el centroide). En aquest gràfic, els individus analitzats es troben a la base del dendrograma i el punt de la seva unió correspon a la distància mètrica/ultramètrica que els separa, que es pot mesurar amb l'escala de distàncies de fusió que es troba a l'esquerra del dendrograma. Com més gran sigui la distància que separa els individus, menor serà la seva similitud química i, a partir d'uns criteris més o menys objectius, es definiran els grups que formen l'estructura del dendrograma i els individus que no s'associen a cap dels grups de la població en estudi. Aquesta funció pren en compte la relació existent entre les diverses variables que formen un grup i si un individu és molt diferent a la resta en quedarà exclòs.

L'anàlisi de Components Principals és una tècnica d'anàlisi multivariant que permet el reconeixement de pautes i relacions entre variables, expressant-se en una matriu de covariàncies. L'objectiu és trobar les components principals (CP_1, CP_2, \dots, CP_D) que són combinacions lineals de les variables originals (X_1, X_2, \dots, X_D) que es descriuen per a tots els individus, segons

$$\begin{aligned} \text{CP}_1 &= a_{11}X_1 + a_{12}X_2 + a_{13}X_3 + \dots + a_{1D}X_D \\ \text{CP}_2 &= a_{21}X_1 + a_{22}X_2 + a_{23}X_3 + \dots + a_{2D}X_D \text{ etc.} \end{aligned}$$

D'aquesta manera, cada component principal ($\text{CP}_1, \dots, \text{CP}_D$) seria la combinació lineal de les diferents composicions químiques transformades en logaritmes de raons (X_1, X_2, \dots, X_D). Aquesta combinació origina un conjunt de noves variables que no estan correlacionades i que formaran les components principals. D'aquesta manera, la primera component principal, CP_1 , recull la major part de la variació que hi ha en el conjunt de dades. La segona, CP_2 , recull la següent major part de la variació i així successivament. Si la correlació és significativa, el nombre de CPs útils serà menor que el nombre de variables originals. Quan una component principal recull la major part de la variació, reduïm la quantitat de dades a tractar, ja que treballarem amb una CP_1 en una dimensió en lloc de treballar en un espai D-dimensional (X_1, \dots, X_D). Les components principals són equivalents a una rotació dels eixos originals, de tal manera que la CP_1 es troba en la direcció de la màxima variació, però mantenint l'angle entre els eixos. La CP_2 es troba en la direcció de la següent variació major i així successivament. Normalment, entre CP_1 i CP_2 recullen la major part de la variació del conjunt de dades. Com a resultat, les dades es poden representar en dues dimensions en lloc de les D originals. A cada CP li correspon un valor que respon a la quantitat de variància del conjunt de dades recollida per aquesta CP. Aquesta variància és compartida per totes les components principals resultants, recollint, la CP_1 més variància que la resta de CPs. A més, s'obté la variància explicada per cada component principal i la proporció acumulada entre components principals. Les variables que més contribueixen a la CP_1 i CP_2 seran les que més variabilitat presenten en el conjunt de dades. Si la variabilitat explicada és important, la distribució que presenten els individus en el gràfic bivariant representat per les CP_1 i CP_2 , proporcionarà una bona aproximació de les distàncies en l'espai D-dimensional. A més, permet definir els grups que conformen la població estudiada i reconèixer les variables que millor discriminen aquestes agrupacions (Aitchison, 1983; Davis, 1986; Baxter, 2006).

L'altre mètode d'anàlisi multivariant utilitzat és l'anàlisi discriminant (AD). Aquest, en comptes d'identificar els grups d'una estructura desconeguda, com hem vist en el mètode aglomeratiu o el de CP, ens proporciona la via per distingir els individus que formen grups prèviament definits, mitjançant criteris independents derivats de les

dades composicionals transformades en logaritmes de raó. L'objectiu és maximitzar les diferències entre els diversos grups definits prèviament per tal d'aconseguir representar les dades en dues dimensions. El punt o eix que separa i divideix les agrupacions és la funció discriminant pròpia de cada cas. Segons la distribució i correlació de les mostres formant grups en el gràfic bivariant, marcarem un eix que passarà pel punt central que divideix les dues agrupacions. Així, aquesta anàlisi calcula la contribució de cada una de les variables considerades a la discriminació dels grups establerts, així com el grau de probabilitat que un individu pertanyi o no a un dels grups definits (Davis, 1986). A més, s'obté, igualment que en l'ACP, el valor propi de cada funció i la contribució de cada una de les variables originals a ella (Davis, 1986; Baxter, 2006).

Els models estadístics teòrics i les tècniques d'agrupament descrites han estat els utilitzats en el tractament estadístiques de les dades composicionals obtingudes mitjançant l'anàlisi química de les àmfores. Tanmateix, existeixen altres models i tècniques d'anàlisi que han estat recollides per Baxter (2001), utilitzades en altres laboratoris arqueomètrics i que l'ús de les quals, en molts casos segons ell, no fa variar en gran mesura els resultats obtinguts mitjançant les tècniques que aquí proposem. La diversitat de mètodes d'agrupament és referida també per Papageorgiou i altres (2001). Independentment de la tècnica d'anàlisi, els grups definits seran considerats grups de referència de cada producció, i vindran expressats a partir de la seva mitjana, com a mesura de tendència central, i de la desviació estàndard, com a mesura de dispersió.