

Integració de mètodes quantitativs als estudis de territori: anàlisi estadística multivariant

Rocío Gómez

rocio.gomezma@campus.uab.cat
Universitat Autònoma de Barcelona

Joan Negre

joan.negre@uab.cat
Becari Predoctoral FPI/UAB (Grup de Recerca OCORDE)
Universitat Autònoma de Barcelona

Resum

L'Arqueologia no és més -ni menys- que una ciència, l'objecte d'estudi de la qual són les conseqüències materials de l'acció humana. Amb la formalització d'aquestes conseqüències a través de les matemàtiques, aconseguim transformar una anàlisi arqueològica en una problemàtica de tipus històrica, a resoldre a través de la reflexió de l'arqueòleg o l'arqueòloga. No serveix per a res quantificar qualsevol tipus de dades sense una metodologia correctament assimilada; la tasca descriptiva és completament insuficient. Amb l'ús de les ciències exactes no pretenem donar-li un fals cos de científicitat a l'Arqueologia, sinó ser conscients que aquestes són l'única eina rigorosa per formalitzar la variabilitat observable al registre arqueològic i, així, plantejar l'estructura causal de la qual prové. En aquest article ens centrarem en com aquests mètodes poden ajudar-nos en les argumentacions sobre les quals construirem els nostres estudis de territori i, a través d'un exemple pràctic, demostrarem la gran diversitat d'informacions que ens poden oferir.

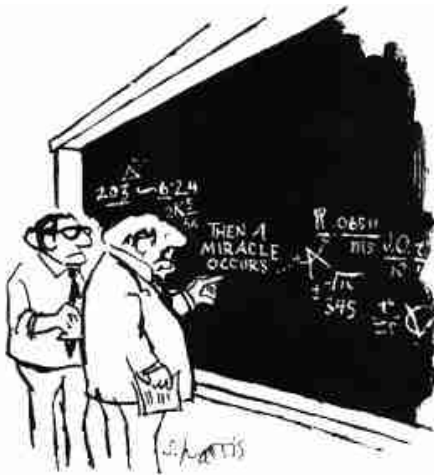
Paraules clau: Arqueologia social, Arqueologia quantitativa, Anàlisi multivariant, Territori, Poblament

Abstract

Archaeology is not more (nor less) than a science, whose objects of study are the material consequences of human action. With the formalization of these consequences through mathematics, we fulfill to turn archaeological analysis into a historical issue which can be answered by the archaeologist. Trying to quantify any possible data without a correct methodology is nonsense; description is not enough. We are not trying to make Archaeology more scientific by using mathematics, but we need them because they are the only rigorous tool to formalize the variability of the archaeological material, in order to suggest the causal structure from where this variability comes. In this article we will focus on how these methods can help us with the arguments on we will base our territorial studies. Besides, we will show the large amount of information that they offer with a practical example.

Key words: Social Archaeology, Quantitative Archaeology, Multivariate analysis, Territory, Settlement

Amb aquest article no intentem il·lustrar un compendi de fórmules, matrius i complicats teoremes amb corol·laris incomprensibles; el que es pretén és presentar la utilitat dels mètodes quantitius per a la investigació arqueològica en general, i més concretament amb algunes aplicacions dins els estudis de territori. Proposarem l'addició a aquests estudis de diversos tipus d'anàlisis quantitatives, des de l'inici del nostre projecte, a l'hora de fer càlculs sobre la viabilitat de la metodologia a emprar, fins al moment de la interpretació crítica del registre material, la variabilitat del qual tan sols podrà ser comprovada a partir de la inferència estadística. L'estadística no serà en cap cas un mètode de demostració d'unes hipòtesis plausibles, sinó més aviat una eina més per a l'argumentació de l'arqueòleg o l'arqueòloga respecte al plantejament d'aquestes (ABELSON, 1998, p. 199 i ss.).



"Crec que hauria de ser una mica més explícit en aquest pas"

Per a la correcta comprensió d'aquest text, i dels objectius que ens proposem, creiem necessari realitzar una succinta definició del que entenem per un estudi de territori, donada l'ambigüitat que pot plantejar aquesta terminologia. Són, de

manera sintètica, documents històrics construïts a força de totes les fonts d'informació disponibles, principalment arqueològiques, però també incloent-hi l'ús crític de la documentació, i en els quals es reflectiran els canvis i dinàmiques socials pròpies d'un espai topogràficament coherent. Aquests tipus de treballs també han estat fomentats per l'anomenada Arqueologia del paisatge (BARCELÓ PERELLÓ, 1988, p. 195), i encara que la denominació d'estudis de paisatge o en extensió és igualment vàlida i pràcticament sinonímica, utilitzant la definició de territori que tot just proposem, creiem més convenient fer servir la nostra nomenclatura.

Atenent als treballs del geògraf Joan Eugeni Sánchez Pérez (1991, p. 15 i ss.), qui defuig tot determinisme mediambiental, es proposa una vertadera relació dialèctica entre els grups humans i el medi físic en què habiten, en tant que tota societat necessita un espai on desenvolupar la seva activitat i obtenir recursos materials, al mateix temps que aquest espai influeix en les transformacions d'aquest grup. L'esmentat autor considera el territori, entès com un espai social, com una variable més en les relacions humanes, ja que tot grup modifica i condiona l'espai en què habita. Així, una de les possibles expressions de canvi social és la modificació de la implantació espacial d'una societat determinada, és a dir, l'organització del poblament. En conclusió, l'espai analitzat per nosaltres és un espai produït, un espai social, atès que és el resultat de l'actuació dels grups humans sobre el medi físic; un medi que estarà condicionat tant per l'esfera econòmica (que controla els processos de producció primaris) com per l'esfera ideològica (que controla la reproducció social) d'aquests grups, la qual cosa ens donarà la possibilitat d'observar la seva estructura social (MENASANCH, 2003)¹. Amb els estudis de territori, i sobretot amb la importància de les dades que una correcta arqueologia pot aportar a aquests, podem replantejar o matisar, a partir de les evidències materials, vells axiomes sobre les dinàmiques socials, suscitats principalment des de l'antropologia, la sociologia o la història més

documental. També hem de ser conscients, però, de les limitacions de l'arqueologia: si bé el seu objecte d'estudi, com ja hem dit, són les conseqüències materials de l'acció social, aquestes són simples mostres del registre material -la part del qual és observable-, que sempre serà incomplet. Així, és més que possible que no siguem capaços de reconstruir una visió sincrònica completa d'una societat; tot i això, hem de poder observar el rastre de materialitat dels processos diacrònics, o dit d'una altra manera, hem de poder detectar les regularitats històriques en la reproducció d'accions col·lectives específiques (BARCELÓ, 2007, p. 14), les quals són el nostre objecte de coneixement².

Causalitat i quantificació en l'arqueologia

En general, l'aplicació de mètodes quantitius de manera generalitzada en l'Arqueologia és un fet que ja duu dècades gestant-se, encara que a l'estat espanyol les mostres d'aquesta aplicació siguin quantitativament -encara que no qualitativament- molt menors. L'existència d'importants òrgans internacionals dedicats a aquesta branca de la nostra disciplina evidencia tant la necessitat d'entrar en contacte amb totes aquestes tècniques, com l'actiu debat que en el centre de la investigació es produeix a nivell mundial respecte a aquestes. El CAA (*Computer Applications and Quantitative Methods in Archaeology*) és l'òrgan internacional més important de coordinació i publicació d'aquests debats, que reuneix tant a arqueòlegs i arqueòlogues, com a membres del món de les matemàtiques i de la informàtica, del qual tenim la sort de comptar amb un membre del seu Comitè de Direcció a la Universitat Autònoma de Barcelona, el professor Dr. Joan Antón Barceló. Les accions d'aquest organisme començaren a la Universitat de Birmingham el 1973, i des de llavors, anualment, s'han reunit per plantejar els principals avenços i mancances d'aquesta disciplina. L'Arqueologia Quantitativa també disposa d'una secció pròpia a la UISPP (*Union Internationale des Sciences Préhistoriques et*

Protohistoriques), un important col·lectiu d'arqueòlegs i arqueòlogues internacional amb més de cent trenta anys d'història. Així, veiem com aquestes tècniques que ara plantegem, ni són noves, ni són dubtosament necessàries, sinó que tenen un important bagatge intel·lectual darrere i un dels debats científics més actuals i actius de les últimes dècades.

El desnivell, però, es palesa durant la formació d'arqueòlegs i arqueòlogues, ja que, al contrari del que ocorre en Geografia, els mètodes quantitius encara són difícils de trobar als diferents programes docents de la nostra llicenciatura. A pesar dels importants manuals publicats ja fa més de vint anys -en algun cas més de trenta- (Doran i Hodson (1975), Orton (1980) o Shennan (1997)), que fins i tot han estat traduïts al castellà també fa més d'una dècada, pocs són els professors que s'aventuren a entrar en aquest món de números, al qual aprengueren, ja fa massa anys, a tómer. El plantejament de qüestions formalment i analíticament més complexes, a més, encara que ja plantejades per a les ciències socials en la dècada dels seixanta, no ha estat fins als anys noranta que ha arribat d'una manera sistematitzada a la docència universitària. Avui dia, per desgràcia, els i les estudiants seguim sense tenir una bona formació en aquestes tècniques durant la llicenciatura, fins i tot a la Universitat Autònoma de Barcelona que, amb les honoroses excepcions de dues assignatures dirigides pel Departament de Prehistòria -l'únic que sembla plantejar-se seriosament la formació metodològica dels futurs arqueòlegs i arqueòlogues-, sembla igualment poc preocupada per l'actualització dels programes docents.

Quant a l'aplicació de les idees de causalitat i quantificació en la recerca, s'ha de ser conscient de les limitacions que l'observació del registre material suposa, tal com hem dit, encara que l'estudi de la variabilitat del mateix ens ajudarà a proposar unes hipòtesis plausibles sobre les causes que l'han convertit en allò que nosaltres estudiem. El punt d'inici de les nostres conseqüències observables estableix l'estructura

de causalitat generadora del mateix registre material, és a dir, la manera com aquest s'ha transformat des d'una matèria primera, cap a un producte social, un ecofacte³ i finalment una resta arqueològica; aquest procés és reflex al mateix temps de canvis en el sistema social artífex de la seva producció. Les matemàtiques són l'únic sistema abstracte de representació de la variabilitat observable en aquest registre, primer pas en la cerca de les seves causes; per tant, saber utilitzar-les per representar i, posteriorment, interpretar les estructures causals que conformen els processos històrics d'una societat són alguns dels principals objectius de l'Arqueologia Quantitativa. És de l'observació del registre arqueològic i dels canvis en les propietats empíriques d'aquesta materialitat (forma, textura, grandària, composició i localització espai-temporal) que podrem determinar la interacció entre diverses accions socials causalment significatives, que seran indicatives de certes dinàmiques socials i de la seva propagació a través d'unes determinades estructures causals (BARCELÓ, 2000). Aquesta variabilitat a què fem referència la mesurarem a través de proves de comparació entre diversos grups de variables i factors, a partir d'uns mètodes estadístics i gràfics concrets, com ara l'elaboració d'histogrames, les mesures de la dispersió dels conjunts o les anàlisis multivariants complexes.

Perquè aquestes mesures de variabilitat tinguin validesa en la formulació d'hipòtesis, serà necessària l'ampliació màxima del mostreig d'observació del registre arqueològic, sempre atents a no contaminar els conjunts amb elements conseqüència de contextos o d'accions diferenciades. És en aquest punt on rau la importància d'un treball de camp minuciós; i de la comparació entre aquests conjunts, extraurem certes regularitats històriques generalitzables a la totalitat del registre. Pel que fa a les relacions entre les diverses variables, estudiarem la possible incidència d'unes amb altres com a factors causals d'unes determinades conseqüències materials. Aquesta causalitat vindrà definida habitualment per relacions complexes, dependents de diverses variables al mateix

temps; i d'esmentada causalitat, la necessitat d'unes tècniques estadístiques més complexes, com l'anàlisi multivariant. En les relacions multidimensionals buscarem correlacions i covariacions entre aquestes variables, establint diversos factors o graus de mesura, útils per a l'investigador o la investigadora en l'ordenació d'aquestes dades, les quals informen del grau de variabilitat dels conjunts d'observacions.

Podem posar un exemple per tenir més clars aquests conceptes: ens trobem davant d'una situació de canvi, descrita per la historiografia simplement com una crisi i un abandonament dels assentaments d'una determinada zona. En primer lloc, observarem i descriurem les variables i registres materials que els defineixen, seguides de la determinació de quins d'ells són abandonats i quins no ho són a partir de la seva estratigrafia. La comparació d'ambdós grups a partir de les seves variables descriptives, correctament ponderades, ens ajudarà a entendre les causes que propiciaren el seu abandonament, com per exemple la diversificació de les produccions com a factor de pes en la seva supervivència, contra l'especialització d'aquells que desaparegueren. En aquest exemple hipotètic, els valors extrems de les variables indicarien una manca de capacitat de reacció davant una situació nova, mentre que uns valors propers al centre de normalitat de les variables mostrarien una major flexibilitat d'aquests grups, i per tant un major índex de supervivència al canvi. El pes relatiu i la incidència de cada criteri en la variabilitat del conjunt de factors descriptius dels jaciments, ens proporcionaran l'estructura causal que ha provocat les conseqüències materials que finalment hem observat sobre el terreny.

L'anàlisi estadística multivariant en l'arqueologia

En termes generals, l'anàlisi multivariant es refereix a totes les tècniques estadístiques que simultàniament analitzen mesures múltiples recollides sobre cada unitat d'anàlisi, tant objectes com de subjectes (MARTÍNEZ, 1999, p. 10).

Encara que alguns autors tan sols utilitzen aquest terme quan la relació de variables estudiades (dependents o independents) segueix una distribució multivariant normal, o quan les variables estan interrelacionades, nosaltres creiem que la mesura, explicació i predicció de les relacions entre variables és més important que el resultat final de l'anàlisi. Quant als objectes d'estudi, les variables que els defineixen poden ser classificades depenent d'una sèrie de factors de classificació; serà important aquesta definició ja que afectarà els tipus d'anàlisi que podrem plantejar sobre aquestes poblacions. Si les definim atenint-nos a si són variables que indiquen les condicions a què exposem els objectes i subjectes d'investigació o si aquestes són explicades per una o més variables, estarem parlant respectivament de variables independents o dependents. En canvi, si posem èmfasi sobre la seva capacitat de ser mesurades en escales mètriques o de no ser-ho, parlarem també respectivament de variables quantitatives/mètriques o variables qualitatives/no mètriques. Trobarem molts altres tipus de classificacions de variables, depenent dels valors que poden adoptar (continues, discretes, binàries/dicotòmiques...) o en funció de les escales i de l'origen, però amb les dues classificacions descrites podrem abordar pràcticament qualsevol tipus d'anàlisi estadística. Les principals àrees o finalitats de l'aplicació d'una anàlisi multivariant són:

Reducció de dades: simplificació de l'estructura del fenomen estudiat, que proporciona un esquema més simple per facilitar la seva interpretació. Parlem per exemple de la creació d'uns factors, resultat de la interacció entre dues o més variables, com podria ser un factor de potencialitat agrícola fruit de les relacions entre el tipus de sòl, la proximitat de recursos hídrics i la distància al nucli de distribució més proper.

Classificació i agrupació: utilitzaríem aquest tipus d'anàlisi per crear grups d'objectes o variables similars entre si, amb la finalitat de crear uns models generals de semblança. Així, podríem fer servir aquesta anàlisi per crear uns grups

excloents de poblament realment vàlids i no basats en criteris de divisió subjectius (assentaments en pendent, d'urbanisme hipodàmic, de fundació castral, etc.). No només classificarem jeràrquicament un seguit d'assentaments (ja hem analitzat la importància que té per nosaltres la plasmació espacial d'una determinada organització social), sinó que advertirem de la importància relativa dels factors que determinen aquesta classificació.

Anàlisi de relacions de dependència entre variables: amb propòsits tant predictius com explicatius, aquest tipus d'anàlisi intenta aprofundir en les relacions de dependència o interdependència de certes variables respecte d'altres. Per exemple, podríem utilitzar aquest tipus d'anàlisi per determinar el pes específic de la variable distància als camps de conreu en l'elecció de l'emplaçament d'un lloc d'hàbitat, o per establir si la relació entre tipus de sòl i tipus de conreu és constant o si existeixen per elecció o per voluntat del grup productor.

Construcció de models i proves d'hipòtesis: encara que les tècniques d'anàlisi multivariant són essencialment descriptives i argumentals, també ens proveeixen d'eines de comprovació d'hipòtesis a través de models complexos basats en poblacions multivariants.

En resum, aquestes àrees d'aplicació, podríem dir que els dos grans tipus d'anàlisi multivariant són aquells que estudien la dependència entre variables i aquells que estudien la interdependència entre aquestes. Entre els mètodes de mesura de la dependència tindriem les tècniques de regressió múltiple, l'anàlisi discriminant i l'anàlisi MANOVA/MANCOVA, que principalment serveixen per estimar diferents tipus de relacions entre variables independents i dependents. Serà l'estudi d'interdependència, però, aquell que centrarà la nostra atenció, atès que ens proporciona una gran varietat de possibilitats en l'anàlisi de múltiples relacions entre variables al mateix temps, molt útil en els estudis de territori. Entre aquestes tècniques destacariem l'anàlisi de conglomerats (*Cluster*

analysis), de components principals (*Principal components analysis*), factorial (*Factor analysis*) o de correspondències (*Correspondence analysis*).

Tècnica	Nº de variable	Tipus de variable	Usos principals
Anàlisi de conglomerats	P	Quantitatives i dicotòmiques	Agrupar els elements de la mostra en grups homogenis en les seves p variables
Escalament multidimensional	P	Quantitatives i dicotòmiques	Explorar les dimensions subjacents en la percepció d'estímuls i preferències
Anàlisi factorial	P	Quantitatives	Reducció del nombre de variables a un menor nombre de factors
Models logarítmic-lineals	P	Qualitatives	Explicar els logaritmes de les freqüències de taules de contingència
Anàlisi de correspondències	2 (simple) o P (múltiple)	Qualitatives	Similituds de files i columnes i relacions entre categories de variables

Figura1: Mètodes d'anàlisi d'interdependència

El procés d'aplicació de l'anàlisi multivariant

Tot seguit presentem un esquema bàsic en l'aplicació de qualsevol anàlisi multivariant dividit en sis fases principals, en les quals es plantejarà, desenvoluparà i argumentarà el problema a través de l'aplicació d'aquestes tècniques (MARTÍNEZ, 1999, p. 78-79):

Fase 1: Definir el problema d'investigació, analitzant els nostres objectius i les hipòtesis de partida, identificant els conceptes i plantejant les preguntes que esperem respondre. L'empirisme a cegues molts cops dona cap resultat, més enllà d'unes relacions entre variables que cap valor donen a la nostra argumentació. També serà el moment de seleccionar les variables que emprarem durant l'anàlisi i la tècnica que més s'adaptarien a aquestes.

Fase 2: Desenvolupar el pla d'anàlisi, lligat a la tècnica que haurem triat, i escollir la grandària del mostreig, tant el d'estimació com el de validació.

Fase 3: Validar uns suposats bàsics per a les mostres, el que es coneix en el món de l'estadística com a *exploració inicial de les dades*. Aquestes tècniques ens ajudaran a determinar la viabilitat de les nostres dades, detectant observacions aïllades (*outliers*), tractant dades perdudes (*missing data*) i comprovant altres premisses necessàries per a l'anàlisi⁴. En set senzills passos podem assegurar-nos que les tècniques que utilitzem seran les idònies, evitarem molts errors en l'aplicació d'aquestes, i aconseguirem uns resultats òptims perquè les nostres argumentacions tinguin una sòlida base. L'ordre de l'*exploració inicial de dades* no és indiferent, ja que els resultats d'un pas afecten el següent.

- Inspecció estadística dels descriptius univariants (valors fora de rang, mitjanes i desviacions típiques, coeficient de variació...)
- Observació de la quantitat i de la distribució de valors perduts i tractament del problema
- Identificació i tractament de les variables

no normals (anàlisi de asimetria i de curtosi, gràfics de probabilitat normal...)

- Identificació i tractament de possibles valors aïllats uni i multivariants
- Examen per mitjà de diagrames de dispersió de la linealitat de les relacions entre variables
- Examen per mitjà dels contrastaments estadístics oportuns de l'homogeneïtat de les variàncies. Si les variables independents són quantitatives poden examinar-se els gràfics de residus
- Avaluació de les variables en referència a la multicolinealitat i singularitat

Fase 4: Estimar el model i el seu ajustament. La majoria de paquets estadístics són comprovats a través d'uns índex que l'investigador o la investigadora haurà de verificar per tal de triar el model d'anàlisi idoni.

Fase 5: Interpretar l'anàlisi resultant, tenint en compte que per a la majoria de tècniques aquest serà algun tipus de variable, una combinació lineal de les puntuacions o una representació en un espai de k dimensions on es situaran les variables originals o els objectes d'estudi.

Fase 6: Validar el model i la comprovació de la seva capacitat perquè sigui generalitzable, per exemple, comparant el model obtingut amb una nova mostra de validació creuada.

A continuació observarem l'aplicació d'aquestes tècniques d'anàlisi multivariant en els estudis del territori, que esperem que serveixi, no tant per a analitzar profundament la praxi d'aquestes tècniques -per a això ja hi ha manuals especialitzats a què fem referència en la bibliografia-, sinó més aviat per a veure com la utilitat d'aquestes no es limita simplement a la creació d'uns models. Ens centrarem en dues de les anàlisis més utilitzades en aquest camp, l'anàlisi de conglomerats i la factorial, les quals

descriurem a través d'alguns exemples. El primer tipus d'anàlisi pertany al grup de tècniques **Q**, és a dir, que estudien les relacions entre objectes o unitats d'anàlisi, mentre que el segon tipus pertany al grup de tècniques **R**, que estudien les relacions entre les variables.

Integració en els estudis de territori: clusters, factors i discriminació

L'Anàlisi de conglomerats reuneix un conjunt de tècniques interessades en la representació d'un reduït conjunt de grups o conglomerats definits per la seva semblança amb una sèrie de característiques o atributs. Aquests grups seran regits per una alta homogeneïtat interna, i una alta heterogeneïtat externa, i encara que les tècniques de classificació poden rebre diversos noms -tècniques Q, mètodes de classificació o agrupació, taxonomia numèrica, etc.-, l'objectiu classificador és comú a tots ells (EVERITT, 1993). En moltes ocasions, aquests tipus d'anàlisi han estat criticats des del món de l'Arqueologia, ja que poden imposar estructures inapropiades o no existents per al conjunt d'unitats d'observació. No hem de caure, però, en l'error que ja hem esmentat d'utilitzar l'estadística com una eina demostrativa, sinó com a eina argumental; ni tampoc podem plantejar un estudi empíric a cegues com ja hem comentat abans; aquestes tècniques superen en mesura de confiança els vells criteris tipològics, i disposen de proves de fiabilitat (lambda de Wilks, delta de Jardine i Sibson, coeficient de correlació cofenètic, etc.) per assegurar al màxim la seva plausibilitat. Tenint en compte aquests conceptes previs i plantejant correctament la investigació, hauríem de ser capaços de localitzar les falsedats intrínseques en l'aplicació de l'anàlisi, i solucionar-les a través d'una correcta interpretació de les dades. La seva utilització en l'anàlisi espacial, a més, ha estat provada amb èxit en diferents casos (BAXTER, 1994, p.148-149; CASTILLO, 1996), tant a nivell microespacial com a nivell macroespacial, utilitzant les coordenades de localització, les variables definitòries dels assentaments, el registre material dels mateixos

per identificar aquests grups, i les seves relacions amb l'espai.

El disseny de l'anàlisi

El primer pas que s'efectua en una anàlisi d'aquest tipus és el de la selecció de variables, així com les seves mesures, que no podrà deslligar-se de la base teòrica de partida ni de les consideracions que se'n deriven. Així, si partim de la base teòrica d'una classificació tenint en compte criteris relatius a la producció, distribució i consum del producte obtingut, haurem de tenir en compte una sèrie de variables que determinaran la creació d'aquests conglomerats. Per a la tècnica de K-mitjanes aquestes dimensions hauran de ser quantitatives, encara que la majoria de tècniques de *clusters* accepten una gran tipologia de mètriques. Un cop elegides aquestes variables, l'investigador o la investigadora haurà de realitzar un seguit de comprovacions per assegurar la correcció dels resultats finals, d'entre els quals, la presència d'observacions aïllades, l'estandardització o ponderació de les variables i finalment l'elecció de la mesura de semblança o no-semblança.

L'anàlisi de conglomerats es veu bastant alterada pels elements singulars aïllats, i en conseqüència és molt aconsellable utilitzar algun tipus de correcció per a aquests, com ara les proves de distància euclidiana o de Mahalanobis. L'estandardització també és una mesura perillosa, ja que pot reduir o modificar les distàncies entre grups d'observacions; els estudis que sobre aquesta qüestió s'han realitzat disten molt encara de la seva homogeneïtat. La reducció de dades, a partir de tècniques com l'anàlisi factorial o de components principals, és en general una opció de certa importància, a causa sobretot de l'excés de pes relatiu que diverses variables correlacionades poden tenir sobre el conjunt d'anàlisi si no es redueixen al factor comú. Finalment, la mesura de semblança és bàsica en la tasca exploratòria de les dades precedent a l'anàlisi final; en resum, no seria més que la mesura de la correspondència entre els objectes. Aquesta pot ser apreciada a partir de tècniques

com les mesures correlacionals (McKEOWN i THOMAS, 1988) -poc utilitzades, per la manca d'incidència en els diversos nivells de les característiques d'aquests objectes-, les mesures de distància -les més utilitzades, en representar aquesta semblança a partir de la proximitat entre les observacions- i les mesures d'associació (EVERITT, 1993) -utilitzades principalment per a variables no mètriques-. La majoria d'aquests procediments poden ser aplicats a partir de programes estadístics informàtics, com ara l'SPSS. Respecte als suposats indicats en la Fase 3 d'investigació (*supra* 3.1), l'anàlisi de conglomerats no és una tècnica d'inferència estadística, sinó un mètode de quantificació estructural d'un conjunt d'observacions, pel que, malgrat tenir propietats matemàtiques, aquestes no són estadístiques; això significa que els suposats de normalitat multivariant i homoscedasticitat, importants per a altres tècniques, no marcaran l'aplicació d'aquesta (MARTÍNEZ, 1999, p. 121).

Solucions i avaluació del model

Hi ha diversos tipus de procediments per a la creació de conglomerats, un cop hem aconseguit la matriu de semblances, que podríem dividir en dos grans grups: els mètodes jeràrquics (aglomeratius o divisius), i els mètodes iteratius de partició.

Procediments jeràrquics: responen a la creació de grups, en forma d'arbre, a partir de la divisió o aglomeració dels objectes des d'una posició inicial. Les tècniques aglomeratives parteixen de la consideració de cada unitat d'observació com un conglomerat, i redueixen en cada pas de l'anàlisi aquests grups a partir de relacions de proximitat o de semblança. El procés contrari, el de divisió, és idèntic, encara que en sentit contrari, i tant l'un com l'altre tenen l'avantatge de poder accedir al nivell immediatament inferior si es creu que el conglomerat creat pot ser erroni o excessivament forçat. Encara que els aglomeratius són més utilitzats, les tècniques per determinar la proximitat d'aquests grups són comuns en ambdós: *single linkage*, basat en la

distància mínima entre objectes; *complet linkage*, basat en la màxima; *average linkage*, que utilitza la distància ponderada de tots els individus entre ells; el *mètode de Ward*, que calcula la distància a partir de la suma de quadrats de totes les variables mesurades; i finalment el *mètode de centroide*, utilitzat a partir de la distància euclidiana entre els seus centroides.

Procediments de partició iterativa: no treballen a partir d'una estructura arbòria, sinó que assignen els objectes un cop establert el nombre de *clusters* amb el que treballarem. Així, es crea una primera partició en els conglomerats indicats, s'assigna cada objecte al conglomerat el centroide del qual s'assembla més, es calculen els nous centroides i es repeteixen els passos centrals fins que els objectes no canvien de conglomerat.

La utilització d'uns o d'altres procediments no és una ciència exacta, en moltes ocasions s'opta per utilitzar els dos i comparar els resultats, amb els quals podrem observar el perquè de les dissimilituds, si n'apareixen. La decisió del nombre de conglomerats també és controvertida, ja que la complexitat de les distribucions de variables impedeix establir una hipòtesi nul·la viable, amb la qual cosa s'opta per realitzar diverses proves amb diferent nombre de *clusters* i així intentar analitzar els resultats que ens proporcionin. La valoració d'ajustament dels resultats dependrà dels procediments utilitzats. Per a les tècniques jeràrquiques, la *correlació cofenètica* és la més utilitzada, determinant l'ajustament entre el dendrograma resultant i el patró de semblances entre les dades. Amb els procediments de partició, s'utilitzen procediments com ANOVA i MANOVA, o l'anàlisi discriminant per examinar la significació de les diferències (MARTÍNEZ, 1999, p. 124; CONOLLY i LAKE, 2004, p. 162-168). Finalment, passàriem a la interpretació de les dades, dels resultats, examinant els components de cada conglomerat i intentant donar una explicació plausible de les causes que condueixen a la seva classificació, i la relació amb la teoria utilitzada de partida.

Aplicació de les tècniques en la investigació

Per estimar el valor d'aquest tipus d'anàlisi en una investigació arqueològica actual, estudiarem el cas de la *Campiña de Jaén*, estudiada per al període comprès entre els segles VIII i X per Juan Carlos Castillo Armenteros (1996). Aquest exemple suposa la integració d'un elevat nombre de factors per permetre una lectura històrica, però sense caure en un detallisme que perdi de vista les característiques principals, en un historicisme estèril. L'autor realitza una anàlisi complexa per a crear una classificació d'assentaments en relació a una sèrie de variables principals, que permetin correlacionar de forma «objectiva» els assentaments, per analitzar l'organització de l'espai. Cal tenir en compte que les classificacions tipològiques poden ser de tres tipus: tradicionals, caracteritzades per l'elecció intuïtiva i qualitativa dels tipus, o més bé dels tipus-ideal, que funcionaran com a fòssil guia per a la resta de mostres; atributius, en què l'anàlisi dels atributs a través d'anàlisis matemàtiques crea uns tipus basats en aquests; i finalment, la que utilitzarem nosaltres, proposada per Doran i Hodson (1975) que consideren el tipus com un grup d'objectes semblants, característica mesurada a partir d'anàlisis quantitatives, tant en la seva mesura com en la seva classificació. Així, i tornant a la definició d'estudi del territori que proposàvem anteriorment, utilitzarem aquestes tècniques per ampliar les argumentacions que podem emprar en el plantejament d'unes hipòtesis versemblants respecte al canvi social reflectit en un determinat territori.

Aquest autor realitza, per al total de jaciments detectats i documentats en aquest territori, una classificació multivariant, amb la qual s'han tingut en compte nou aspectes principals que recullen les característiques físiques, econòmiques i estratègiques quantificables de cada jaciment. Existeix també la possibilitat de combinar aquesta classificació amb variables no-mètriques, com el lloc d'emplaçament, les estructures més abundants o els materials més abundants. Els factors que s'han tingut en compte en aquest

cas (MOLINOS, SERRANO, RÍSQUEZ i MONTILLA, 1994) s'estructuren en dos blocs:

territorials: aquells obtinguts de l'estudi de les característiques físiques de l'assentament, relleu, geologia, edafologia, etc. Són:

- grandària: gran dificultat per quantificar-la, atès que els jaciments no solen ser excavats en la seva totalitat, i que aquesta normalment és determinada per l'àrea de dispersió dels materials. A més, els urbans en cap cas poden ser mesurats per la seva situació de destrucció, només poden ser localitzats a partir del topònim, fonts escrites, o una petita zona excavada en un solar. Hi ha diverses opcions per determinar la grandària del jaciment, entre elles, canvis en la coloració de les terres, prospeccions GPR i dispersió dels materials.
- altitud absoluta sobre el nivell del mar
- altitud relativa de l'assentament sobre els seu territori d'explotació, determinat a partir d'un *site catchment analysis* (WHEATLEY i GILLINGS, 2002, p. 159-162; CONOLLY i LAKE, 2006, p. 213-225)
- pendent mitjana de l'àrea d'explotació potencialitat agrícola: presenta algunes dificultats, derivades principalment del canvi del valor i ús del sòl actual respecte al pretèrit. S'han de tenir molt en compte els estudis paleoambientals abans d'intentar extreure aquest índex. Esmentat factor està lligat a diverses variables climatològiques, geogràfiques, geològiques i antròpiques, encara que creiem important fer una aproximació plausible en aquest índex. Per a realitzar aquesta operació podem utilitzar una anàlisi factorial, o també una superposició raster en GIS de les diverses variables utilitzades per mesurar aquest índex. L'anàlisi de potencialitat agrícola estarà basat en la superfície de *site catchment* calculada per a cada assentament.

- territori en zona d'alta fertilitat: amb aquest factor s'incrementa el valor de certes zones de màxima potencialitat agrícola.

històrics: a partir de l'elecció i l'establiment dels assentaments sobre l'espai. Giren sobretot al voltant del potencial estratègic de cada assentament.

- distància amb veí més proper: bastant important quant a la distribució espacial del territori i significativa per establir comparacions entre les estructures de poblament de la mateixa o d'una altra zona. En tot cas, és un factor a prendre amb precaució, ja que no s'han localitzat tots els jaciments existents, encara que pot ser un mostreig acceptable, depenent de cada cas i de les proves de significació aplicades.
- àrea de control visual, a partir d'una anàlisi de visibilitat a través d'un GIS, tenint en compte tots els problemes que per a aquest tipus d'anàlisi ja s'han esmentat a la bibliografia (WHEATLEY i GILLINGS, 2002, p. 201-214; CONOLLY i LAKE, 2006, p. 225-233).
- concentració visual: nombre d'assentaments visibles des de cadascun d'ells.

A l'hora de realitzar l'anàlisi, caldrà observar els resultats obtinguts a partir de la matriu de dades de totes les variables conjuntament i a partir del desglossament de les variables en territorials o històriques, per comprovar l'ajustament dels resultats. En el cas dels jaciments urbans, per no modificar el valor de l'anàlisi, realitzarem la seva exploració per separat, eliminant la variable grandària de la seva matriu d'observacions. El resultat d'aplicar-ho, per exemple al cas de la *Campiña de Jaén* entre els segles VIII-X (CASTILLO, 1996), és la creació de deu tipus d'assentaments, per a una mostra de vuitanta-set jaciments per a les variables territorials i de dotze per a les històriques. Aplicant sobre aquesta classificació *cluster* dels nostres assentaments una anàlisi factorial dels resultats, trobem unes

solucions altament correlacionades, amb un factor mínim del 51,7%, però en la majoria dels casos, per damunt del 85%. A continuació, utilitzarem l'anàlisi discriminant per a verificar i contrastar els resultats d'ambdues anàlisis. Els tipus predefinitos seran els obtinguts a l'anàlisi de conglomerats, per tal d'observar si en funció d'aquests, queden correctament discriminats, i quines correccions es poden realitzar. També utilitzarem en aquesta discriminació estadística les variables més determinants en la classificació de les observacions, obtingudes a partir de l'anterior anàlisi factorial, que ens ajudaran a ponderar els resultats.

Finalment, serà necessària la comparació de les anàlisis de les variables territorials i històriques del assentaments extraurbans d'aquells que estan sota les actuals ciutats, per tal de veure si també segueixen uns patrons comuns, i determinar els conglomerats finals resultants. A partir d'aquí, comença la tasca més important, que és la d'interpretar i analitzar cada *cluster*, al mateix temps que localitzem la distribució geogràfica d'aquests per a analitzar-los respecte als seus veïns més propers. L'últim pas, doncs, seria integrar aquesta argumentació obtinguda a partir de l'anàlisi multivariant complexa, en la nostra proposta d'interpretació plausible dels fenòmens i dinàmiques socials d'una determinada societat. En aquest cas, els resultats obtinguts serien els següents:

Cluster I combinat: el més nombrós, amb el 37,93% de les observacions, amb altituds mitjanes/baixes, un pendent reduït i una potencialitat agrícola mitjana, amb poca zona d'alta fertilitat. L'àrea de control visual tindria valors mitjans/baixos i amb molt pocs veïns (1-2) en aquesta zona.

Cluster II combinat: amb un 16,09% de les observacions, destaca per una altitud mitjana, amb un pendent mitjà/alt, una potencialitat agrícola mitjana i una escassa àrea d'alta fertilitat. L'àrea de control visual és molt limitada per la topografia, i és un tipus d'assentament molt difós pel territori.

Cluster III combinat: amb un 3,45% és un grup poc representat, amb unes altituds mitjanes/baixes, poc pendent, potencialitat agrícola mitjana i accés a zones d'alta fertilitat.

Cluster IV combinat: amb un 1,15% de la mostra, queda representat únicament per un jaciment, diferenciat de la resta dels de la seva zona per un alt pendent, baixa potencialitat agrícola i un accés limitat a zones d'alta fertilitat.

Cluster V combinat: amb un 2,30%, respon a dos jaciments caracteritzats per una altura i un pendent baixos amb alts índex de potencialitat agrícola i distàncies importants amb els seus veïns, sense una àrea de control visual gaire gran.

Cluster VI combinat: amb un 1,15%, també representat únicament per un jaciment, es diferencia pel seu baix índex d'altitud relativa i pels seus importants recursos agrícoles en situar-se, a diferència dels seus veïns, en un fons de vall amb importants recursos hídrics.

Cluster VII combinat: s'ubica en el curs del riu Guadalquivir, constituït per dos assentaments, amb un alt potencial agrícola però amb una altitud relativa i pendent més elevats, per la seva posició de control estratègica.

Cluster VIII combinat: representa el 10,34% del total, destaca per les seves altituds i pendents mitjans, que destaca dins la tònica general. La seva potencialitat agrícola és mitjana/alta, i amb zones d'alta fertilitat importants.

Cluster IX combinat: amb un 12,64% de les observacions, queda agrupat gràcies a les variables de distància respecte als veïns més propers i pel seu nivell de control visual, més alts que els de tota la resta. Les seves altituds i pendents no són molt alts, i disposen d'una alta potencialitat agrícola.

Cluster X combinat: diferenciat de la resta de grups, amb un 4,6% de la mostra, aquest conglomerat es caracteritza per una potencialitat

agrícola baixa, un accés a zones d'alta fertilitat reduït i una àrea de control visual molt baixa.

Cluster XI combinat: també format per un sol jaciment, destaca pel seu alt potencial agrícola i per la seva important àrea de control visual respecte als seus veïns.

Cluster XII combinat: Amb un 4,6% dels jaciments, situat principalment en la Vall del Guadalquivir, en terrasses relativament poc elevades, controla un ampli territori, amb una molt alta accessibilitat a zones de gran fertilitat.

Cluster XIII combinat: finalment, amb un 2,3%, és l'assentament amb els pendents més alts i les potencialitats agrícoles més baixes de tota la mostra.

Amb aquestes dades seria important, a continuació, analitzar aquesta classificació a partir de la seva relació amb el territori, amb els assentaments que els envolten, i sobretot, amb les dinàmiques de grup que poden seguir. Així, en els casos amb unes poblacions baixes, podem intentar entendre el tipus de decisions de grup, depenent del tipus de jaciment, que ha marcat la seva localització, i la seva funció dins l'elecció general d'aquella societat. Amb quatre grups o tendències generals per sobre del 10% dels casos estudiats, la importància de determinar la lògica dels petits conglomerats es converteix en una prioritat en l'anàlisi de l'organització espacial del territori. Així hem construït, definitivament, una nova eina per a l'argumentació del nostre estudi, una eina científica -i per tant, rigorosa i objectiva-, d'explicació de les estructures causals responsables d'aquella materialitat social que l'arqueologia ha estudiat.

Cap a una arqueologia social i científica

Hem arribat a la fi de la nostra exposició, i creiem que hem aportat una primera aproximació a l'ús de les eines estadístiques -d'una ínfima part d'elles- que poden ajudar en l'argumentació de qualsevol estudi de territori rigorós. Esperem que

tot aquell que hagi llegit aquest text pugui haver observat la necessitat d'utilitzar aquestes eines, i les immenses possibilitats que ens ofereixen al món de l'Arqueologia en concret. Perdre la por a les matemàtiques i no relegar-les al calaix de l'oblit és l'opció que cada investigador o investigadora ha de prendre en algun moment de la seva formació; ens agradaria pensar que la seva demostrada utilitat serà raó de pes suficient com per aconseguir-ho. Com en aquest mateix número proposem, la científicitat de l'Arqueologia no és una quimera al cap d'uns pocs, sinó l'únic camí lògic a seguir per a construir una argumentació contrastable -i per suposat, oberta a matisacions- que enriqueixi la capacitat crítica de la societat a què ens devem, i per a la qual investiguem. L'Arqueologia serà una eina d'utilitat social en el moment en què el coneixement que aquesta aportari a tota i totes les societats pugui ser utilitzat per entendre la realitat, primer pas necessari per canviar-la (FERNÁNDEZ, 2005).

El mètode científic és l'únic sistema vàlid que tenim perquè les conclusions a què arribem siguin plausibles i rigoroses; d'aquí que l'Arqueologia Quantitativa sigui l'única via possible en l'estudi del registre material observable, en tant que com a matèria, té propietats mètriques no subjectives que són la base del seu estudi. Per tant, treure-li importància a aquests principis i abordar el seu estudi a partir de variables o etiquetes qualitatives fruit de l'observació subjectiva de l'arqueòleg o l'arqueòloga, ens allunya d'aquesta pretesa i rigorosa científicitat de les nostres interpretacions. A partir d'aquestes propostes, l'Arqueologia serà capaç d'iniciar una relació dialèctica amb la societat respecte als límits i el futur en la nostra percepció del canvi social. Seguim creient que aquest debat, proposat ja per la nostra anterior generació, és imprescindible; el proveïment a la societat d'aquesta eina, l'Arqueologia, és la nostra petita contribució, amb dades contrastables, perquè la gent pugui extreure les seves pròpies conclusions respecte als problemes del sistema social en què vivim i les solucions que li poden donar. La decisió d'utilitzar-la ja no depèn de nosaltres. Estarem aquí per quantificar el seu grau d'èxit.

Agraïments

Agraïm la inestimable ajuda rebuda tant en la revisió d'aquest article, com en els consells bibliogràfics, orientatius i exemplificatius respecte al món de l'Arqueologia Quantitativa del professor Joan Antón Barceló, del Departament de

Prehistòria de la UAB, responsable també del Laboratori d'Arqueologia Quantitativa i d'Aplicacions Informàtiques. En qualsevol cas, els possibles errors en el text són solament responsabilitat dels autors del mateix.

Amb el suport de la Universitat Autònoma de Barcelona.

Notes

¹ En aquest punt, hem d'agrair enormement les recomanacions bibliogràfiques i orientatives de la professora Maria Juana López Medina, de la Universitat d'Almeria, qui ens ajudat amb les seves apreciacions.

² Respecte a la diferenciació que fem entre objecte d'estudi i objecte de coneixement al nostre escrit, s'hauria d'indicar l'arbitrarietat en l'ús d'aquests tipus de conceptes. Des d'un altre punt de vista, es podria dir que l'estudi de l'objecte ens duu al seu propi coneixement, així com l'estudi de la societat ens conduirà al coneixement d'aquesta. La utilització d'aquesta diferenciació indicaria que l'estudi d'un objecte es transforma en coneixement de la societat, que és el primer sempre «objecte de desconeixement», i queda la segona sempre «sense estudiar». Per tant, seguint aquest punt de vista,

l'objecte tant d'estudi com de coneixement de l'Arqueologia seria el procés o processos de canvis socials, a través de les conseqüències materials visibles, que a la vegada són també estudiades i conegudes per l'arqueòleg i l'arqueòloga.

³ Entenem per ecofacte aquell artefacte que està subjecte a transformacions biològiques, i no socials. És el pas intermedi que proposem com a necessari entre un producte social i una resta arqueològica (CAMARÓS, 2007, p.74).

⁴ Per raons d'espai no podem analitzar detalladament totes les tècniques de l'exploració inicial de dades, pel que recomanem al lector que, en primer lloc, llegeixi l'excel·lent primer volum de la col·lecció de manuals d'estadística per a Arqueologia de J.A. Barceló (2007), i posteriorment l'obra de R. Martínez (1999, p. 44-77) citada a la bibliografia.

Bibliografia

ABELSON, R.P. *La estadística razonada: reglas y principios*. Barcelona: Paidós, 1998

BARCELÓ PERELLÓ, M. "La Arqueología extensiva y el estudio de la creación del espacio rural". *Arqueología medieval. En las afueras del "medievalismo"*. Barcelona: Crítica, 1988

BARCELÓ, J.A. *Arqueología y estadística: Introducción al estudio de la variabilidad de las evidencias arqueológicas*. Bellaterra: Servei de publicacions de la Universitat Autònoma de Barcelona, 2007

BARCELÓ, J.A., BRIZ, I. i VILA, A. (ed.). *New*

techniques for old times: CAA98. Oxford: British Archaeological Reports International, Series 757, 1999

BAXTER, M.J. *Exploratory multivariate analysis in Archaeology*. Edimburgh: Edinburgh University Press, 1994

BAXTER, M.J. *Statistics in Archaeology*. Londres: Arnold, 2003

CASTILLO, J.C. *Introducción arqueológica a un proceso histórico. El poblamiento humano emiral en la campiña de Jaen*. Tesi doctoral inèdita [microforma]. Jaén: Servicio de Publicaciones e

Intercambio Científico de la Universidad de Jaén, 1996

CONOLLY, J. i **LAKE, M.** *Geographical Information Systems in Archaeology*. Cambridge: Cambridge University Press, 2006

DORAN, J. i **HODSON, F.** *Mathematics and Computers in Archaeology*. Edimburgh: Edimburgh University Press, 1975

EBERT, D. *Expanding the selection of tools for spatial analysis: geostatistics and the Als feildwalking data*. Southampton: University of Southampton, 1998

EVERITT, B.S. *Cluster analysis*. Londres: Arnold, 1993

FERNÁNDEZ, V. *Una Arqueología crítica*. Barcelona: Crítica, 2005

HAINING, R. *Spatial Data Analysis: Theory and Practice*. Cambridge: Cambridge University Press, 2003.

HILLIER, B. i **HANSON, J.** *The social logic of space*. Cambridge: Cambridge University Press, 1984

MARTÍNEZ, R. *El análisis multivariante en la Investigación científica*. Madrid: Cuadernos de Estadística, La Muralla, 1999

MENASANCH, M. *Secuencias de cambio social en una región mediterránea. Análisis arqueológico de la depresión de Vera (Almería) entre los siglos V y XI*. Oxford: B.A.R. Int.Ser. 1132, 2003

MENASANCH, M. *Los "poblados de altura": centros de los nuevos espacios sociales en el sudeste peninsular (siglos V-VIII)*. VI Reunió d'Arqueologia Cristiana Hispànica: les ciutats tardoantigues d'Hispania : cristianització i topografia. València: 8, 9 i 10 de maig de 2003. Any de publicació 2005: p. 375-384

McKEOWN, B. i **THOMAS, D.** *Q methodology*. Beverly Hills: Sage, 1988

MOLINOS, M., SERRANO, J.C., RÍSQUEZ, C. i **MONTILLA, S.** *Un problema de fronteras en la periferia de Tartessos: Las Calañas de Marmolejo*, Jaén: Universidad de Jaén, 1994

ROGERSON, P.A. *Statistical Methods for Geography*. Londres: Sage, 2001

ORTON, C. *Mathematics in Archaeology*. Londres: Collins, 1980.

SANCHEZ-PÉREZ, J.E. *Espacio, economía y sociedad* Barcelona: Siglo XXI, 1991

SHENNAN, S. *Quantifying Archaeology*. 2a edició. Edimburgh: Edimburgh University Press, 1997

TRIGGER, B.G. *Historia del pensamiento arqueológico*. Barcelona: Crítica, 1992

WHEATLEY, D.W. i **GILLINGS, M.** *Spatial Technology and Archaeology: the archaeological applications of GIS*. Londres: Taylor & Francis, 2002

Articles de revistes

BAXTER, M.J. *Methodological issues in the study of assemblage diversity, American Antiquity*. 66 (2001): p. 715-725

BAXTER, M.J., BEARDAH, C. i **WRIGHT, R.** *Some archaeological applications of kernel density estimates. Journal of Archaeological Science*. 24 (4) (1997): p. 347-354

BAXTER, M.J. i **COOL, H.** *Exploring Romano-British finds assemblages. Oxford Journal of Archaeology*. 21 (4) (2002): p. 365-384

CAMARÓS, E. *El procés de transformació de la matèria primera en resta arqueozoològica en les societats caçadores-recol·lectores a Estrat Crític*. 1 (2007): p. 69-75

Article de revista digital

BARCELÓ, J.A. *De la "inutilitat" del passat a la "comoditat" del present. La rendibilitat de les ciències socials en el segle XXI. Digit-HUM Revista digital d'humanitats* (1999): Universitat Oberta de Catalunya

Webgrafia

BARCELÓ, J.A. *La inutilidad de la Arqueología*. (1996). En línia:
<http://antalya.uab.es/prehistoria/Barcelo/inutilidadDeArqueologia.html>
(30 de març de 2008)

BARCELÓ, J.A. *Causalidad en Arqueología y Ciencias sociales*. (2000). En línia:
<http://antalya.uab.cat/prehistoria/Barcelo/CausalidadArqCieSoc.html>
(30 de març de 2008)