

## UN USAGE DES TESTS NON-PARAMÉTRIQUES EN ARCHÉOLOGIE: L'ÉTUDE D'UN DÉPÔT D'ARCHIVES À DÉLOS

La maison de Skardhana à Délos (GD 59D) renfermait un grand nombre de documents lors de sa destruction par un incendie en 69 av. J.-C. D'eux il ne reste que les pastilles d'argile qui les fermaient. Sur chaque pastille, on observe des empreintes, généralement en nombre multiple, et dont chacune est le résultat de l'impression d'une matrice (Figs. 1-2). L'exploitation des données produites par ces pastilles est primordiale car les autres sources d'informations fournissent peu de renseignements. Les documents scellés étaient avant tout des actes d'ordre privé, dont la teneur nous échappe. Une description complète de cette trouvaille est réalisée par M.-F. BOUSSAC (1993).

Le nombre total de pastilles trouvées lors des fouilles de 1974-75 et en 1987 est évalué à ca 16000 pastilles d'argile; le nombre d'empreintes devrait avoisiner 26/27000 numéros et l'ensemble des matrices lisibles se monter à ca 14000. Pour la présente étude, seul le matériel recueilli en 1974-75 dont nous disposons en totalité est étudié. Il comporte 1487 cachets sur lesquels sont apposées 3435 empreintes. L'analyse visuelle des empreintes a permis de répertorier 2225 matrices. Une première étude statistique descriptive réalisée par Y. AUDA et M.-F. BOUSSAC (1996) pose la problématique archéologique et fournit les premiers résultats numériques. Dans la présente étude, l'analyse est complétée par la conception et la mise en œuvre de tests inférentiels adaptés aux questions posées par l'archéologue.

Rappelons la forme sous laquelle se présentent les données. Pour chaque cachet, l'archéologue a identifié les matrices qui ont servi à l'impression des empreintes. Considérons pour fixer les idées un exemple constitué par un extrait choisi dans le catalogue 1975.

Soient sept cachets sur lesquels sont relevées les empreintes qui résultent de l'impression des matrices suivantes :

cachet 1 :	71	1837	1435	1533
cachet 2 :	71	1435	1533	1837
cachet 3 :	71	850		
cachet 4 :	71	1533		
cachet 5 :	1332	1952	190	
cachet 6 :	1332	1533		
cachet 7 :	414	156		

Dans nos analyses, les empreintes trop abîmées pour permettre l'identification de la matrice ne sont pas prises en compte. Dans une première approche, cette simplification ne devrait pas gêner la mise en évidence des structures.

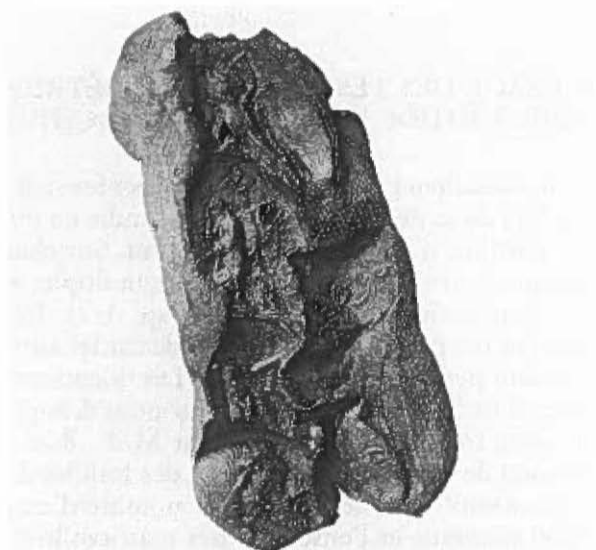


Fig. 1 – Pastille à empreintes multiple.

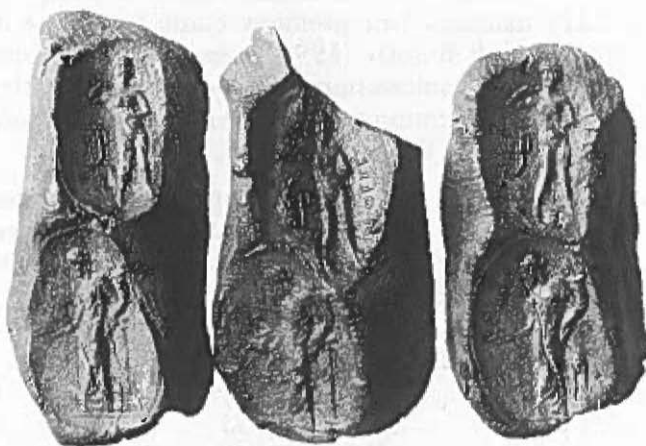


Fig. 2 – Récurrence des mêmes matrices sur plusieurs pastilles.

Rappelons les statistiques de base développées pour décrire ce type de données. On établit la liste des couples de matrices associées, c'est à dire imprimées sur le même cachet : (71,1837), (71,1435), (71,1533), (1837,1435), (1837,1533), (1435,1533), (71,1435), (71,1533), (71,1837), (1435,1533), ((1435,1837), (1533,1837), (71,850), (71,1533), (1332,1952), (1332,190), (1952,190), (1332,1533), (414, 156).

Cette liste permet de construire le tableau symétrique suivant:

	71	156	190	414	850	1332	1435	1533	1837	1952
71					1		2	3	2	
156				1						
190						1				1
414		1								
850	1									
1332			1					1		1
1435	2							2	2	
1533	3					1	2		2	
1837	2						2	2		
1952			1			1				

On peut aussi pour chaque matrice calculer les statistiques d'associations suivantes:

matrice	nombre d'empreintes	nombre d'associations	distribution du nombre associations		
			1	2	3
156	1	1	1	2	1
190	1	2	1	0	0
414	1	1	2	0	0
850	1	1	1	0	0
1332	2	3	1	0	0
1435	2	6	3	0	0
1533	4	8	0	3	0
1837	2	6	1	2	1
1952	1	2	0	3	0
total	19	38	2	0	0
			12	10	2

Le nombre d'empreintes exprime la fréquence d'usage d'une matrice. Le nombre d'associations est fonction de la participation de cette matrice à des cachets à empreintes multiples. Plus un cachet porte d'empreintes, plus le nombre d'associations des matrices imprimées sur ce cachet sera grand. La distribution du nombre d'associations renseigne sur les associations préférentielles de matrice. Par exemple la matrice 71 est retrouvée trois fois avec une autre matrice, la 1533.

La représentation graphique adaptée à ce type de données consiste à matérialiser chaque association de matrices par un trait. Lorsque deux matrices sont associées à plusieurs reprises, un nombre de traits égal à leur nombre d'associations les relie.

Le schéma ci-dessous illustre notre exemple:

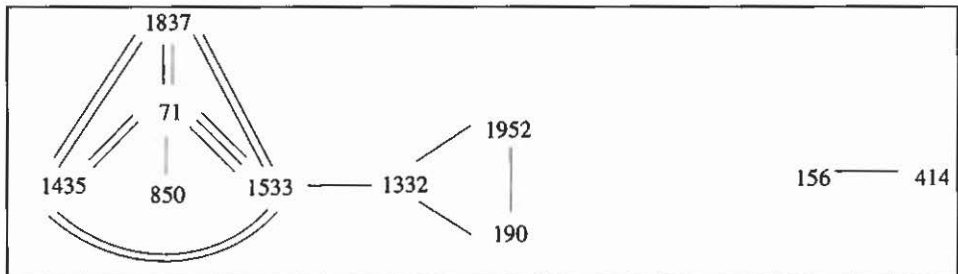


Fig. 3 - Représentation graphique des associations de matrices pour l'exemple extrait des données de la campagne 1975.

Décrivons le mode de lecture de ce schéma. La matrice 71 est présente sur quatre des six cachets, ce qui explique sa position centrale sur le dessin. Par contre, à l'inverse, la matrice 850 n'est présente que sur un cachet. Le trait qui la relie à la matrice 71 indique qu'elle est imprimée sur un cachet qui porte aussi une empreinte de la matrice 71.

Sur ce schéma, l'importance numérique des liaisons multiples renforce l'hypothèse émise par l'archéologue de l'existence de liaison préférentielle entre matrices. Cette observation pourrait s'expliquer par le fait que deux usagers qui ont participé à la constitution d'un document sont en relation et auront donc tendance à répéter l'impression de leur sceau sur un même cachet. Mais nous ne pouvons en apporter la preuve sans réaliser de tests statistiques.

La particularité des données ne permet pas de choisir une méthode parmi les tests usuels. Elle demande de développer des procédures inférentielles propres aux données constituées par les sceaux de Délos. Des méthodes non-paramétriques sont utilisées. Elles sont fondées sur la comparaison d'un critère observé par rapport à une distribution de ce même critère calculé sur des jeux de données qui sont générés par ordinateur.

Notre mise en œuvre et notre conduite des tests s'inspirent beaucoup d'informations disponibles sur le WEB du laboratoire "Math Sciences Computing Center" de l'Université de Washington dont l'URL est <http://www.ms.washington.edu/old-course-pages/stat390-summer96/>.

Nous commencerons par décrire ces tests pour l'exemple précédent réduit à sept cachets avant de les appliquer à la totalité des données recueillies lors de la campagne 1975. Ces tests essayeront de cerner l'existence de relations préférentielles entre usagers de matrice et leur influence sur la constitution du nombre et de la taille des groupes d'affaires organisés autour de ces liaisons. Compte tenu de ces objectifs, le nombre de cachets, le nombre de matrices, le nombre d'empreintes imprimées sur chacun des cachets et le nom-

bre d'empreintes auquel a donné lieu chacune des matrices apparaissent comme des paramètres fixés. Ces contraintes conduisent à présenter les données sous la forme de deux tableaux disjonctifs complets. Le premier tableau est obtenu par le croisement empreinte par cachet. Le deuxième tableau est obtenu par le croisement empreinte par matrice. Reprenons notre exemple précédent pour illustrer cette mise en forme des données.

	cachets						matrices											
	1	2	3	4	5	6	7	71	156	190	414	850	1332	1435	1533	1837	1952	
1							1											
1																	1	
1													1					
1															1			
	1						1											
	1												1					
	1														1			
		1																1
		1																
		1																
			1				1											
			1									1						
				1														
				1														
					1								1					
					1													
						1												
						1												
							1											
								1										
									1									
										1								
											1							
												1						
													1					
														1				
															1			
																1		
																	1	
																		1
totaux	4	4	2	2	3	2	2	4	1	1	1	1	2	2	4	2	1	

La première ligne de ces deux tableaux signifie que la première empreinte parmi les dix-neuf étudiées est imprimée par la matrice 71 sur le cachet 1. Ce mode de construction implique pour ces deux tableaux des marges ligne (somme de tous les éléments de chacune des lignes) égales à un vecteur rempli de 1. La marge colonne (somme de tous les éléments de chacune des colonnes) du premier tableau correspond au nombre d'empreintes imprimées sur chacun des cachets. La marge colonne du deuxième tableau correspond au nombre d'empreintes auquel a donné lieu chacune des matrices.

Générer des jeux de données sans modifier les marges colonne pour respecter les contraintes précédemment définies s'effectue donc par permutation des lignes d'un seul de ces deux tableaux. Il est ensuite possible de calculer un critère, fonction d'une hypothèse à tester, pour chacun des jeux de données obtenus par permutation. L'ensemble de ces calculs construit une

distribution de référence qui permettra d'évaluer le degré de vraisemblance du critère observé. Ce dernier est calculé sur les données qui ne sont pas permutées. Comme il existe, si l'on ne considère que l'extrait de la campagne 1975 réduit à sept cachets imprimés par 19 empreintes,  $19!$  (approximativement égal à  $10^{26}$ ) permutations possibles, les calculs nécessaires à l'établissement de la distribution de référence ne sont réalisés qu'à partir de 10000 permutations tirées aléatoirement par un programme informatique.

Toute la difficulté de l'application d'une méthode inférentielle pour cet exemple est de choisir une hypothèse et donc un critère adapté à la question que se pose l'archéologue. La simple observation des pastilles d'argile laisse entrevoir que certaines matrices sont souvent associées par la présence simultanée de leurs empreintes sur les cachets. A partir de cette constatation, l'analyse est poursuivie par l'examen consécutif des problèmes archéologiques formulés avec précision et le test des hypothèses associées. La progression dans la compréhension de ces données archéologiques, en particulier les liens entre les usagers de matrices, est ainsi assurée. En pratique, nous avons choisi de poser le problème le plus simplement possible, puis de lui associer deux hypothèses. La première, l'hypothèse notée  $H_0$ , exprimera une idée contraire à celle de l'archéologue. La deuxième alternative, notée  $H_A$ , est en accord avec l'avis de l'archéologue. Le rejet de  $H_0$  entérinera  $H_A$  et confirmera le point de vue de l'archéologue. Au contraire, l'acceptation de  $H_0$  conduira l'archéologue à réviser sa position.

#### *Problème archéologique n° 1*

Quand deux usagers impriment leur matrice sur un même cachet, une impression conjointe de ces matrices sur un autre cachet est favorisée.

- $H_0$ : le nombre d'associations uniques est supérieur (ou égal) à celui attendu par le simple fait du hasard.
- $H_A$ : le nombre d'associations uniques est inférieur à celui attendu par le simple fait du hasard.

Pour l'exemple réduit à sept cachets, le nombre d'associations uniques observé est 12. Si l'hypothèse  $H_0$  est vraie, quel est le degré de vraisemblance pour que seulement 12 associations uniques soient observées? Pour le savoir, nous avons écrit un programme qui a généré 10000 permutations des lignes du tableau croisé empreinte par cachet, et qui a calculé après chaque permutation le nombre d'associations uniques. Pour l'exemple réduit à sept cachets, on observe 27 fois une valeur inférieure à la valeur observé 12. Il y a donc 27 chances sur 10000 d'observer un nombre d'associations uniques inférieur à 12. C'est extrêmement peu, et il est très peu probable, si le choix des matrices se fait selon les lois du hasard, d'observer un nombre d'associations uniques aussi faible que 12. L'hypothèse  $H_0$  est rejetée pour l'exemple réduit à sept cachets avec un risque de première espèce proche de 0,0027. Cette technique est appliquée à l'ensemble des données recueillies lors de la campagne

1974-75. Aucun des jeux de données générés ne possède un nombre d'associations simples inférieur à la valeur observée égale à 3956\*. L'hypothèse est donc rejetée avec un risque de première espèce si faible que nous n'avons pu l'estimer. Il ne peut être nul car il y a au moins 1/3435! chance d'observer un nombre d'associations uniques plus grand (ou égal) à 3956. En effet, l'observation réalisée sur les données archéologiques est égale à ce nombre.

### *Problème archéologique n° 2*

Des relations très préférentielles existent entre les usagers les plus importants. L'importance des usagers est estimée par le nombre d'empreintes imprimées par leur matrice.

- $H_0$ : le plus grand nombre d'associations entre deux matrices est inférieur (ou égal) à celui attendu par le simple fait du hasard.
- $H_A$ : le plus grand nombre d'associations entre deux matrices est supérieur à celui attendu par le simple fait du hasard.

Pour l'exemple réduit à sept cachets, le plus grand nombre d'associations observé est 3. Pour l'ensemble des données recueillies lors de la campagne 1974-75, sa valeur est de 6. Aucun des jeux de données générés pour l'exemple ou pour l'intégralité des données ne génère une valeur supérieure. L'hypothèse  $H_0$  est rejetée avec un risque de première espèce si faible que nous n'avons pu l'estimer.

Ces deux tests suffisent à prouver l'association préférentielle des matrices ce qui indiquerait l'existence de petits groupes d'"affaires" dont les membres sont les usagers des matrices. Un groupe est virtuellement défini par l'ensemble des matrices dont au moins une empreinte est apposée sur un cachet où figure l'empreinte d'une autre matrice de ce groupe. La compréhension de cette définition est facilitée par une représentation graphique des relations entre matrice reportées dans la Fig. 4. Un groupe correspond à un ensemble des matrices reliées par des traits qui matérialisent leur impression sur un ou plusieurs cachets communs. Deux membres d'un groupe ne participent pas forcément à des "affaires" communes, mais ils sont en contact par l'intermédiaire d'un ou plusieurs membres tiers.

Essayons de mieux cerner les caractéristiques de ces groupes par le test de nouvelles hypothèses.

### *Problème archéologique n° 3*

L'existence d'associations préférentielles entre matrices répartit les matrices en groupes.

- $H_0$ : le nombre de groupes de matrices est inférieur (ou égal) à celui attendu par le simple fait du hasard.

\* Ce chiffre est différent de celui fourni par Y. AJUDA et M.-F. BOUSSAC (1996) car le présent calcul intègre les matrices en un seul exemplaire.

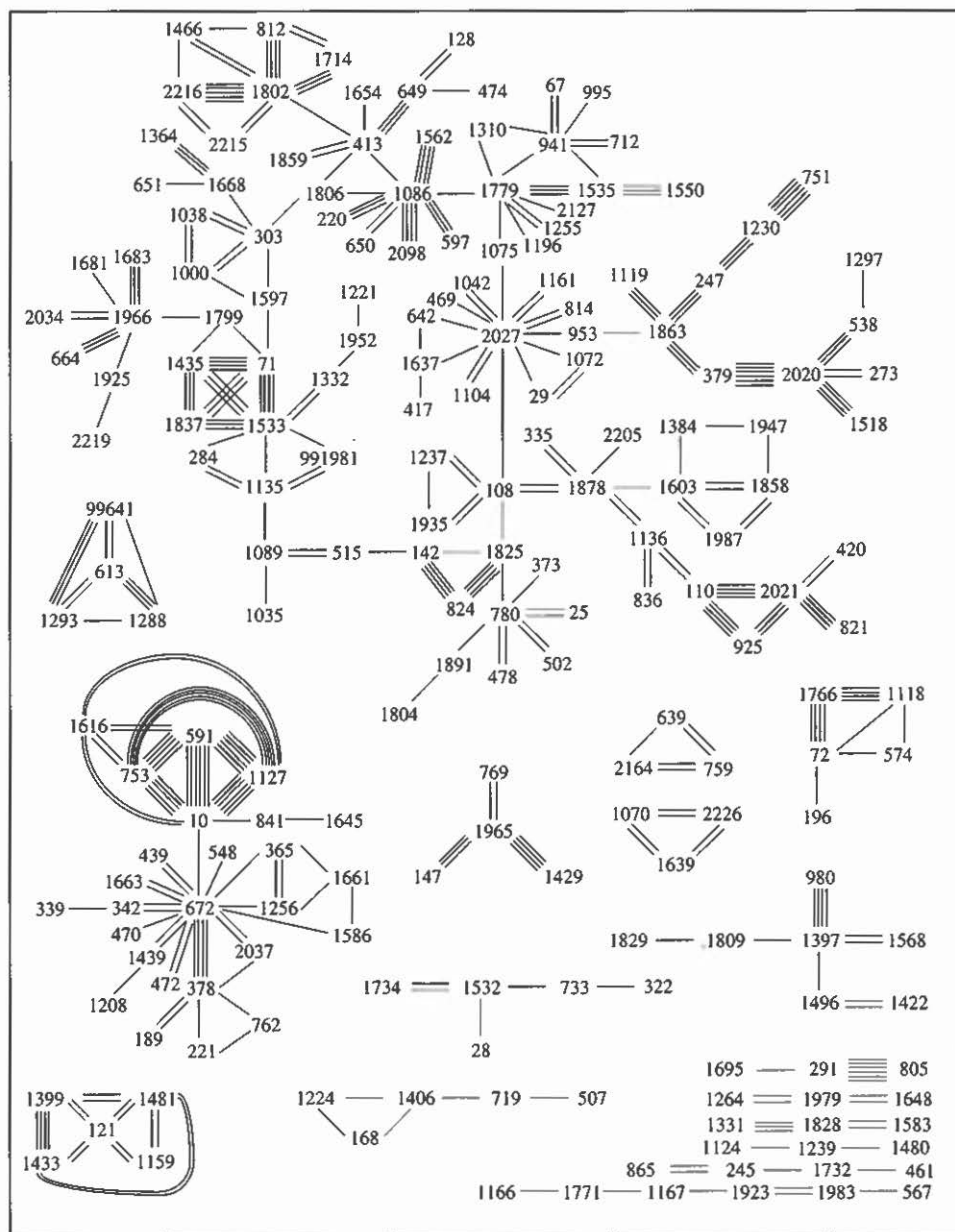


Fig. 4 - Représentation graphique des associations de matrices pour l'échantillon recueilli lors de la campagne 1974-75. Seules les matrices qui participent à des groupes composés de plus de deux matrices sont représentées. Les autres se répartissent en quatre types de groupes avec les occurrences suivantes: 12 A — B; 17 A ≡ B; 5 A ≡ B; 2 A ≡ B.



- $H_A$ : le nombre de groupes de matrices est supérieur à celui attendu par le simple fait du hasard.

Les sept cachets se répartissent en deux groupes. L'ensemble des données en compte 676. Pour l'exemple réduit à sept cachets comme pour l'intégralité des données, les dix mille générations de données ont toujours produit dans tous les jeux de données permutés un nombre de groupes inférieur. L'hypothèse est donc rejetée avec un risque de première espèce si faible que nous n'avons pu l'estimer.

Ce test conforte les résultats précédents. L'existence d'associations préférentielles entre matrices augmente le nombre de liaisons multiples et diminue les liaisons uniques. Ce mode d'association structure fortement les relations entre les usagers des matrices, en particulier en diminuant la possibilité de liens entre deux matrices. Il conduit à la création de nombreux groupes qui n'ont pas de contact entre eux. La représentation graphique des associations de matrice pour un des jeux de données générés serait beaucoup plus touffue que celle de la Fig. 4. De nombreux liens simples reliaient les matrices.

Vérifions que cette structuration en groupes se répercute au niveau de la taille des groupes.

- $H_0$ : le nombre moyen de matrices par groupe est supérieur (ou égal) à celui attendu par le simple fait du hasard.
- $H_A$ : le nombre moyen de matrices par groupe est inférieur à celui attendu par le simple fait du hasard.

Les groupes sont composés en moyenne de 5 matrices, chiffre qui est toujours inférieur à ceux obtenus par calcul sur les jeux de données générés. Cette hypothèse est donc rejetée avec un risque de première espèce si faible que nous n'avons pu l'estimer. Testons l'hypothèse corollaire.

- $H_0$ : le plus grand nombre de matrices qui compose un groupe est inférieur (ou égal) à celui attendu par le simple fait du hasard.
- $H_A$ : le plus grand nombre de matrices qui compose un groupe est supérieur à celui attendu par le simple fait du hasard.

Aucun jeu de données générés ne présente une valeur supérieure à 964, nombre de matrices qui compose le plus grand groupe observé. Cette hypothèse est également rejetée avec un risque de première espèce si faible que nous n'avons pu l'estimer.

Sur l'ensemble des données recueillies pendant la campagne 1974-75, l'existence d'associations préférentielles qui conduisent au cloisonnement des relations entre usagers de matrices par groupe est définitivement prouvée par les tests inférentiels non-paramétriques par permutation. Nous n'oublions pas que ce travail ne porte que sur une partie du matériel recueilli, et que cette partie ne constitue elle-même qu'une fraction des documents conservés dans ces archives. Mais l'étude de cachets supplémentaires ne devrait pas

modifier nos résultats. Le fait qu'un document puisse être scellé par plusieurs cachets ne peut pas non plus changer nos conclusions dans la mesure où les structures mises en évidence sont extrêmement fortes. Certains groupes isolés pourraient peut-être se rattacher aux grands groupes majeurs. L'occurrence des matrices augmenterait sans conteste. Quelques nouvelles matrices seraient découvertes. Mais nous ne pensons pas que ces modifications puissent bouleverser la structure du mode d'association de matrices révélé par nos analyses. Les associations préférentielles au niveau de petits groupes composés le plus souvent de trois à sept empreintes différentes subsisteront. Une hiérarchie forte dirigée par un ou deux usagers des matrices est aussi définitivement exclue.

Ces données de l'analyse numérique confirment les hypothèses formulées précédemment par M.-F. BOUSSAC (1993) sur le ou les propriétaire(s) : en excluant toute structure hiérarchique, elles dénie(n)t à n'importe quel usager un rôle institutionnel. Le grand nombre d'associations attribuées à certains s'expliquerait par la variété et l'importance de leurs affaires, non par leur fonction. L'hypothèse qui fait du ou des propriétaire(s) de la maison des brasseurs d'affaires, probablement banquiers, s'en trouverait renforcée : outre leurs propres papiers d'affaires, ils conserveraient en dépôt, selon un usage attique réintroduit sous la seconde domination athénienne, des actes privés conclus par d'autres et qui n'étaient pas enregistrés. Ils serviraient de dépositaires habituels de contrats, sans pour autant être des professionnels rémunérés pour leur tâche.

YVES AUDA

Maison de l'Orient Méditerranéen  
Lyon

MARIE-FRANÇOISE BOUSSAC  
UMR 9969, Institut F. Courby

#### BIBLIOGRAPHIE

- BOUSSAC M.-F. 1993, *Archives personnelles à Délos*, «Comptes rendus de l'Académie des inscriptions et belles-lettres», 677-693.  
AUDA Y., BOUSSAC M.-F. 1996 (1997), *Étude statistique d'un dépôt d'archives à Délos*, «Bulletin de Correspondence Hellénique» Suppl. 29, 511-525.

#### ABSTRACT

Non-parametrical tests have been used to check the validity of some hypothesis made on archaeological material (sealings on papyri kept in a Delian archive destroyed by a fire in 69 BC). Although these tests deal only with a part of the corpus, the validity of the results should not be questioned.