مجلة جامعة تشرين للدراسات والبحوث العلمية - سلسلة العلوم الاقتصادية المجلد (20) العدد (1) 1998
Tishreen University Journal for Studies and Scientific Research –Economical Science Series Vol. (20) No (1) 1998

مقدمة حول التحليل الإحصائي للبيانات بطريقة المكونات الأساسية ACP والتحليل العاملي التقابلي AFC ملامح منهجية وطرائقية

الدكتور محمود طيوب*

(قبل للنشر في 1996/8/3)

❑ ملخص ❑

يهدف هذا البحث إلى توضيح الأسس والمبادئ التي تستند إليها طريقة التحليل الإحصائي متعدد المتغيرات، لاسيما طريقة المكونات الأساسية *ACP*، والتحليل العاملي التقابلي *AFC*، وشروط تطبيق كل منهما والفروقات بينهما. وتشكل الطريقتان في وقتنا الحاضر الأسلوب الإحصائي الأكثر انتشاراً، ولاسيما بعد ما يسمى بثورة الانفورماتيك، وتوسيع استخدام الحواسيب ذوات الاستطاعات العالية. وتنتمي الطريقتان إلى الإحصاء الوصفي، وتهدفان إلى عرض وتمثيل المشاهدات المرتبطة بعضها ببعضها الآخر، بواسطة عدة متغيرات، في فضاء محدود الاتجاهات بدلاً من عرضها في فضاء متعدد الاتجاهات، شريطة ألا يؤدي ذلك إلى فقدان بعض المعلومات، أي المحافظة قدر المستطاع على البنى الهيكلية الموجودة بين بيانات الجدول الإحصائي المدروس، إضافة لتوضيح كيفية الحسابات من خلال المثالين التطبيقيين.

* مدرس في قسم الإحصاء – كلية الاقتصاد – جامعة تشرين – اللاذقية – سورية.

مجلة جامعة تشرين للدراسات والبحوث العلمية - سلسلة العلوم الاقتصادية المجلد (20) العدد (1) 1998
Tishreen University Journal for Studies and Scientific Research –Economical Science Series Vol. (20) No (1) 1998

Introduction á l'analyse statistique des données par les méthodes ACP-AFC aspects statistiques et méthodologiques

Dr. M. TAYOUB*

(Accepté le 3/8/1996)

❐ RÉSUMÉ ❐

L'analyse en composantes Principales (ACP) et l'Analyse Factorielle des correspondances (AFC) apportent une solution intéressante aux problèmes Posés. Toutes deux relèvent de la statistique descriptive. Elles ont pour but de représenter un ensemble d'observations d'un certain nombre de variables appartenant à un espace de dimension élevée dans un espace de dimension plus faible avec minimum de Perte d'informations c.á.d. de telle sorte que les distances entre les unités statistiques ou entre les variables soient le mieux possible sauvegardées.

* Enseignant au Département de Statistique, Faculté d'Economie, Université de Tichrine, Lattaquié, Syrie.

1- مقدمة Introduction:

تعتبر طرق التحليل الإحصائي "متعدد المتغيرات" المتعلقة بالبيانات الإحصائية مجموعة من التقنيات الحديثة، التي تسمح بوصف، واختصار، وتبويب وتصنيف المشاهدات التي تتصف بعدة متغيرات، وتسمح كذلك بوصف وتوضيح البنى الهيكلية لهذه البيانات (أو العلاقات الموجودة بين هذه البيانات). رياضياً، تستند هذه التقنيات إلى الطرق الجبرية، وهي من حيث المبدأ لا تتطلب مقارنة بالطرق الكلاسيكية أي موديل احتمالي، وتعتبر طرق تحليل البيانات وصفاً للاستدلال الإحصائي، الذي يسمح بحسب (Benzecri, et al. 1973) بطرح افتراضات أو موديل إحصائي قابل للتحقق أدرج (Pielou , 1975) هذه التقنيات الإحصائية كأساس لكثير من الدراسات البيئية التي ترتكز في البحث على أجوبة للمشكلات المطروحة دون أي تصور سابق.

تاريخياً: استخدمت تقنيات التحليل "متعدد المتغيرات" في بداية القرن العشرين (Spearman, 1904; Pearson, 1901) وفي نهاية الثلاثينات وضعت معظم الأسس النظرية الرياضية (Bartlett, 1933; Hotelling, 1936; Anderson, 1958) . إلا أن ضعف مقدرة الحاسبات الإلكترونية، وعدم كفايتها أخر استخدام هذه التقنيات حتى بداية عقد الخمسينات، ولاسيما (Rao, 1952) إن تصميم التقنيات الإلكترونية المتطورة وانتشار بنوك المعطيات، سمحا للباحثين بتطوير طرق ومنهجية معالجة بياناتهم بصورة أكثر دقة وشمولية، ولاسيما استخدام طرق التحليل العاملي Analyse factorielle بواسطة Benzecri بداية عقد الستينات من هذا القرن طريقة التحليل العاملي التقابلي AFC Analyse Factorielles des Correspondences (Benzecri, 1980).

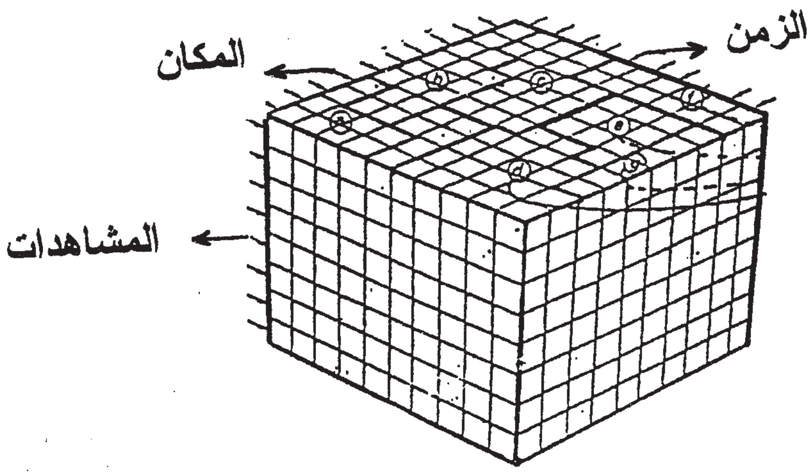
2- الهدف من البحث L'objectif:

انطلاقاً من أهمية هذه التقنيات في معالجة البيانات الإحصائية، رأينا ضرورة إعداد هذه الدراسة، بهدف إبراز خصائص، وميزات، واستخدام، وشروط تطبيق، هذه التقنيات، إضافة لذلك، طرق حسابها، ولاسيما طريقة التحليل العاملي التقابلي AFC، وطريقة المكونات الأساسية ACP، والفروق فيهما، ومجالات تطبيقاتهما.

3- الجداول الإحصائية Tableaux Statistiques:

إن جدول تحليل البيانات عبارة عن مصفوفة أبعادها $n \times m$، تحدد مجموعة الأفراد والمشاهدات -الأفراد- والتغيرات المدروسة. تعتبر مرحلة إعداد الجدول الإحصائي، استعداداً للتحليل، من المراحل الأساسية، وتعكس بدقة طبيعة الظواهر

المدروسة، وانسجامها مع مسألة البحث. ويتطلب إعداد الجدول هنا أسلوباً يتلاءم مع تقنيات التحليل الإحصائي، فمثلاً يتطلب التحليل بطريقة ACP جدولاً خالياً من القيم المعدومة (صفر)، ضمن هذا المفهوم، يرى (Jambus, 1973) أن الجدول الإحصائي متعلق بعدة عوامل أهمها: الهدف، وطبيعة البيانات، والتقنيات الإحصائية المستخدمة...الخ انظر الشكل (1):



الشكل (1) يبين جدول بيانات إحصائية ثلاثي الأبعاد

تسمح الجداول الإحصائية عموماً، ولاسيما المستخدمة في الدراسات البيئية والظواهر الطبيعية الأخرى متعددة المتغيرات، بتحليل البيانات وفق عدة محاور زمانية ومكانية. لقد اعتبر (Legendre & Legendre, 1984) بشكل أساسي أن هذه المحاور مرتبطة بالمتغيرات: الزمن × النوع × المكان.

هذا، ويمكن تحليل البيانات البيئية وفق عدة نماذج منها:

أولاً: دراسة معاملات التشابه Similarité، ويرمز لها بـP: يسمح هذا النموذج بإجراء عدة مقارنات بين الظواهر التي تقسم إلى حالتين (غائب / موجود؛ نعم أو لا...الخ)، وذلك بحساب معاملات الاقتران فيما بينها.

ثانياً: دراسة معاملات التباعد Distance ويرمز لها بـQ: تمثل ظاهرة معاكسة للحالة الأولى، فقيمتها العظمى تساوي الواحد الصحيح عندما تكون الظاهرتان مختلفتين اختلافاً كلياً، وقيمتها الدنيا تساوي الصفر عندما تكون الظاهرتان متشابهتين تشابهاً كلياً، يسمح هذا النموذج بإجراء مقارنات بين محطات توصف بوجود عدة أفراد من ظواهر معينة خلال فترة زمنية

معينة.

ثالثاً: دراسة معاملات الارتباط، Corrélation ويرمز لها بـR: يسمح هذا النموذج بدراسة العلاقات الكامنة بين مختلف الظواهر المدروسة والعوامل المؤثرة في ذلك، باستخدام معامل الارتباط الخطي /لبيرسون/ أو معامل الارتباط الرتبي -/لسبيرمان/ في حالة الظواهر النوعية. بمعنى آخر، دراسة معاملات الارتباط بين وجود بعض الأنواع والعوامل البيئية المختلفة خلال فترة زمنية معينة.

تعتبر تقنيات التحليل العاملي أكثر التقنيات الإحصائية لتحليل البيانات البيئية استخداماً، لما تتمتع به من إمكانيات لدراسة وتوضيح طبيعة هذه العلاقات والعوامل المؤثرة فيها (Escofier et Pages , 1988).

انطلاقاً من أهمية هذه التقنيات ولاسيما بعد شيوع استخدام الحواسب، لابد من إلقاء الضوء على هذه التقنيات، خصوصاً، طريقة التحليل العاملي التقابلي AFC، وطريقة المكونات الأساسية ACP.

1- التحليل بطريقة المكونات الأساسية :ACP

لإعطاء فكرة عن آلية الحساب والتحليل، بالمثال التالي:

- ليكن لدينا الجدول (1) مؤلفاً من ثلاثة متغيرات × ستة مشاهدات:

جدول (1): جدول بيانات 6 × 3 (مثال فرضي):

المتغيرات				
	i	L	I	E
المشاهدات	1	5	2.5	1
	2	3	1	0.7
	3	2.5	2	1
	4	4	3	2.6
	5	4.5	3.2	2.5
	6	1.5	1	0.8

- نحدد الأوساط الحسابية والانحرافات المعيارية:

المتغيرات	الأوساط الحسابية	الانحراف المعياري
L	3.417	1.2047
I	2.117	0.8764
E	1.433	0.7972

• نحسب الارتباط بين المتغيرات، ويعطى بالمصفوفة التالية:

	L	I	E
L	1		
I	0.791	1	
E	0.532	0.858	1

*- دراسة القيم الخاصة والمناحي الخاصة:(**)

القيم الخاصة Valeurs Propres لمصفوفة بيانات A عبارة عن حل المعادلة التالية: $|A-\lambda I|=0$. أما المناحي الخاصة /العوامل/ Vecteur Propre فتحدد بواسطة الصيغة التالية: $|A-\lambda_i I|u_i=0$.

دراسة القيم الخاصة والمناحي (العوامل) الخاصة:

القيم الخاصة	التباين الكلي ٪	التباين التجميعي ٪
$\lambda_1 = 2.4622$	82.07%	82.07%
$\lambda_2 = 0.4709$	15.77%	97.77%
$\lambda_3 = 0.0669$	2.23%	100%

القيمة λ_1 وλ_2 من القيم الخاصة تشكل أكثر من 97٪ من التباين الكلي، لذلك سندرس الآن توزع المتغيرات ضمن الفضاء الشعاعي للعاملين I وII.

(**) ملاحظة: لمزيد من المعلومات لحساب القيم والعوامل راجع المصفوفات.

* المناحي الخاصة للقيمة الخاصة λ_1	* المناحي الخاصة المقابلة للقيمة الخاصة λ_2
0.5421	0.7530
0.6236	-0.0631
0.5632	-0.6550

والمناحي الخاصة (عبارة عن معاملات المتغيرات المركزية المختصرة المعيارية للمعادلة الخطية للمحاور الرئيسية).

***- تمثيل المتحولات Représentation des variables:**

تحسب إحداثيات المتحولات من العلاقة التالية: $a_i = \alpha_1\sqrt{\lambda_i}$

$0.6236.\sqrt{2.4322} = 0.8526$ = المحور الأول: المتحول I.

$0.7530.\sqrt{0.4709} = 0.5167$ = المحور الثاني.

المحور الأول: المتحول II. = $0.6236.\sqrt{2.4622} = 0.9785$

المحور الثاني. = $-0.0631.\sqrt{0.4709} = -0.0433$

المحور الأول: المتحول III. = $0.5632.\sqrt{2.4622} = 0.8837$

المحور الثاني. = $-0.6550.\sqrt{0.4709} = -0.4495$

وترتب النتائج على الشكل التالي:

المحور الثاني F_2	المحور الأول F_1	المتحولات
0.5167	0.8506	L
-0.0433	0.9785	I
-0.4495	0.8837	e

- **تمثيل المشاهدات Présentation des Sujets:**

لحساب إحداثيات المشاهدات الملاحظة i المحور الأفقي يمثل العامل h ويحسب من العلاقة التالية:

$$a_{i(h)} = \sum_{j=1}^{P} x_{ij} . S_j$$

حيث X_{ij}: البيانات المختصرة.

S_j: مكونة j للعامل الخاص h.

P: عدد العوامل المعتبرة.

الملاحظة (الفرد الأول) الإحداثيات على المحور I:

$$= \frac{5-3.4167}{1.3197}.0.5421 + \frac{0.5-2.1167}{0.9600}.0.6236 + \frac{1-1.4333}{0.8733}.0.5632 = 0.6199$$

الإحداثيات على المحور II:

$$= \frac{5-3.4167}{1.3197}.0.4530 + \frac{2.5-2.1167}{0.9600}.(-0.0631) + \frac{1-1.4333}{0.8733}.(-0.6550) = 1.2032$$

$$= \frac{X_{ij} - \overline{X}_j}{S_i}.V_P.\lambda_1 + ... +$$

وهكذا بالنسبة لباقي المشاهدات، وترتب النتائج وفق الجدول (2):

جدول (2): يبين إحداثيات المشاهدات على المحورين الأول والثاني لـــACP.

المشاهدات	المحور الأول F_1	المحور الثاني F_2
1	0.6199	1.2032
2	-1.3694	0.3857
3	-0.7318	-0.1904
4	1.5658	-0.6002
5	1.8366	-0.2531
6	-1.9212	-0.5452

- **دائرة معامل الارتباط Cercle de Corrélation:**

ضمن فضاء المتحولات، إذا اعتبرنا كل المكونات (وهي $P = 3$) يكون لدينا $\sum \alpha_i^2 = 1$، فكل متحول يتحدد بنقطة على محيط دائرة الارتباط، وبحسب الموضع المقابل للمتحول L على الشكل التالي:

$$OL = \sqrt{a_i^2} = \sqrt{0.8506^2 + 0.5167^2} = 0.9904 \approx 1$$

نستنتج أن تمثيل المتحول L على مخطط المحورين $II \times I$ يعتبر جيداً لأن قيمته قريبة من الواحد، وأن موضع نقاط الانتشار (للمتحولات) بالنسبة لمحيط الدائرة ستعطي فكرة عن نوعية تمثيل البيانات، إضافة لذلك (نقاط المتحولات) ستكون قريبة من محيط الدائرة، وبالتالي فإن تمثيل البيانات أكثر دقة وأهمية (Cailliez et Pages, 1976).

- **حساب Cos^2 للوحدات الإحصائية بالنسبة للمحاور العاملية:**

تقديراً لنوعية التمثيل للوحدات الإحصائية، سندرس ما يسمى بمربع تجيب الزاوية θ_j بالنسبة لكل محور P، وليكن لدينا من أجل كل المحاور: $\sum_{j=1}^{P} \cos^2 \theta_j = 1$.

إذا حسبنا $\sum \cos^2 \theta$ للفضاء المأخوذ بعين الاعتبار (العوامل المرغوب تمثيلها)، فإن نوعية التمثيل أفضل بكثير، والتي ستكون قريبة من الواحد، إذا: $\theta_{i(h)}$: تمثيل زاوية الشعاع المقابل للوحدة الإحصائية i بالنسبة للعامل h إذاً نلاحظ أن:

$$\cos^2 \theta_{i(h)} = \frac{\alpha_i^2}{\sum_{j=1}^{P} u_{ij}^2}$$

حيث إن: α_i: المحور الأفقي للملاحظة i بالنسبة للمحور h.

u_{ij}: البيانات المختصرة للملاحظة i.

مثال: لنحسب $Cos^2\theta$ للملاحظة الأولى بالنسبة للمحور الأول:

$$\cos^2\theta_{1(F_1)} = \frac{0.6199^2}{\left(\frac{5-3.4167}{1.3197}\right)^2 + \left(\frac{2.5-2.1167}{0.9600}\right)^2 + \left(\frac{1-1.4333}{0.8737}\right)^2} = 0.2082$$

وهكذا بالنسبة لباقي الملاحظات: كما في الجدول (3):

جدول(3): يبين قيم $\cos^2$ للمشاهدات على المحورين الأول والثاني ACP.

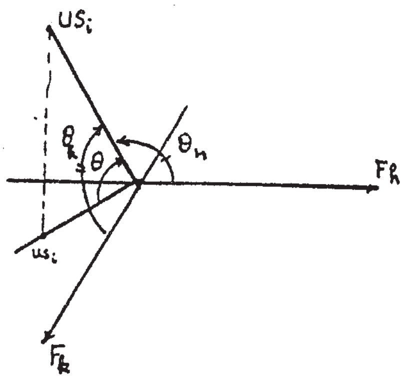
$\cos^2\theta$	المحور الأول F_1	المحور الثاني F_2	Totale
1	0.2082	0.7847	0.9929
2	0.8690	0.0689	0.9379
3	0.7203	0.0488	0.7691
4	0.8673	0.1274	0.9947
5	0.9808	0.0186	0.9994
6	0.9254	0.0745	0.9399

ومنه نستنتج تمثيلاً ممتازاً ضمن فضاء العوامل المأخوذة.

- **حساب $\cos^2$ للوحدات الإحصائية عن المخطط (F_h, F_k):**

$$\cos^2\theta(F_h, F_k) = \cos^2\theta_h + cps^2\theta_k$$

تمثيل الوحدة الإحصائية i على مسطح المخطط F_h, F_k يكون أفضل بحساب مربعات تجيب الزوايا $\cos^2\theta_{(F_h,F_k)}$، والتي تقرب من الواحد الصحيح، كما هو موضح بالشكل (2):



الشكل (2) يبين تمثيل وحدة إحصائية على المخطط الهندسي.

- **حساب مشاركة المتحولات x_i على المحور F_k كنسبة مئوية:**

إن مشاركة المتحول x_i المستند على العامل المميز K المقابل له كما يلي:

$$C_{(x_i,F_k)} = 100.a_{ik}^2 \Big/ \sqrt{\sum_{i=1}^{P} a_k^2} = 100.a_{ik}^2$$

لأن $\sum a_{ik}^2 = 1$

مثال: مشاركة المتحولات على المحور الأول:

$100 \times 0.5421^2 = 29.38\%$ مشاركة المتحول L.

$100 \times 0.6236^2 = 38.88\%$ مشاركة المتحول I.

$100 \times 0.5632^2 = 31.71\%$ مشاركة المتحول e.

مشاركة المتحولات على المحور الثاني:

L: $100 \times 0.7530^2 = 56.7\%$
I: $100 \times (-0.0631)^2 = 0.398\%$
e: $100 \times (-0.6550)^2 = 42.9\%$

- **حساب مشاركة المشاهدات على المحور F_k كنسبة مئوية:**

وتحسب بواسطة الصيغة التالية:

$$C_{(i,F_k)} = 100.\beta_{ik}^2 \Big/ \sqrt{\sum_{i=1}^{n} \beta_{ik}^2}$$
$$= 100.\beta_{ik}^2 / (n-1).\lambda_k$$

حيث إن: β_{ik}: إحداثية الملاحظة i عن المحور F_k.

n: عدد المشاهدات.

λ_k: القيمة الخاصة المقابلة.

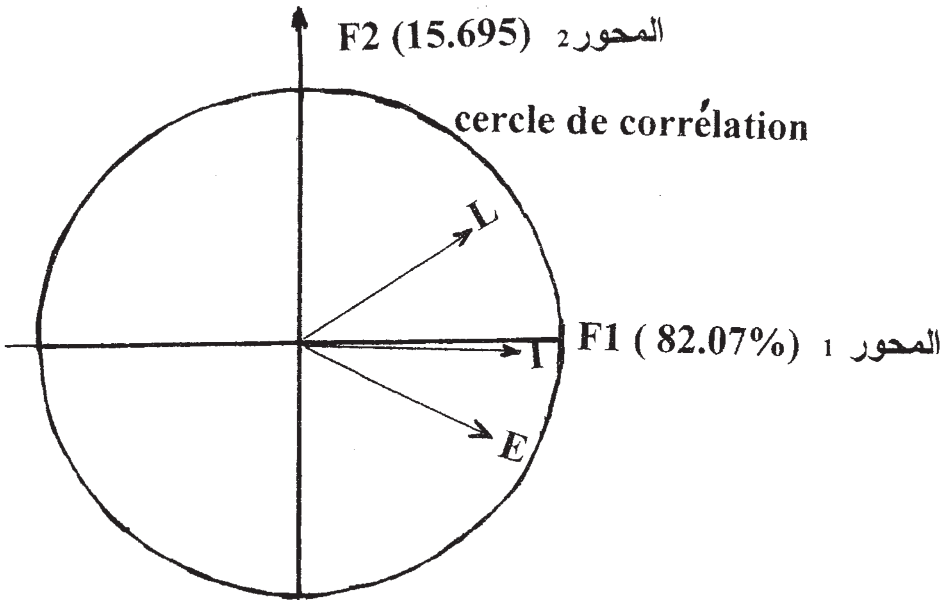
- مشاركة المشاهدات على المحور الأول كنسبة مئوية على النحو التالي:

1:	100 . 0.6499^2 / (6-1) . 2.4622	= 3.14%
2:	100 . $(-1.3694)^2$ / (6-1) . 2.4622	= 15.23%
3:	. . . .	= 4.35%
4:	. . . .	= 19.91%
5:	. . . .	= 27.40%
6:	. . . .	= 29.98%

- حساب مشاركة المشاهدات على المحور الثاني على النحو التالي:

1:	100 . 1.2032^2 / (6-1) . 0.4709	= 61.48%
2:	100 . 0.3857^2 / (6-1) . 0.4709	= 6.32%
3:	. . . .	= 4.35%
4:	. . . .	= 19.91%
5:	. . . .	= 27.40%
6:	100 . $(-0.5452)^2$ / (6-6) . 0.4709	= 12.62%

• التمثيل البياني لدائرة معامل الارتباط:



الشكل (3): تمثيل المتغيرات على المسطح العاملي للمحورين الأول والثاني لـــــACP (دائرة الارتباط).

تمثيل المشاهدات على المحور 1×2: لـACP:

الشكل (4): التمثيل البياني للمشاهدات على المحورين الأول والثاني لـــــACP.

2- التحليل العاملي التقابلي AFC:

ليكن لدينا توزيع /5/ أنواع من الحشرات يرمز لها بـE,D,C,B,A في أربعة أوساط بيئية مختلفة، هي: وسط مائي، يرمز له بـEAU، وسط أشنيات SPH، وسط تربة مع طحالب MSO، وسط صخور مملوءة بالطحالب MRO موزعة وفقاً للجدول (4):

جدول (4): يبين توزع أربعة أنواع من الطحالب في أربعة أوساط مائية (مثال فرضي)

الأوساط / الأنواع	وسط مائي EAU	وسط به أشنيات SPH	طحالب على تربة MSO	طحالب وصخور MRO	المجموع الكلي TOTAL
A	1	2	3	4	10
B	0	2	3	4	و
C	5	3	2	0	10
D	4	3	2	1	10
E	5	5	0	0	10
$\sum$	15	15	10	9	49

* حساب جدول التكرارات النسبية:

	EAU	SPH	MSO	MRO	P_i
A	1/49	2/49	3/49	4/49	10/49
B	0	2/49	3/49	4/49	9/49
C	5/49	3/49	2/49	0	10/49
D	4/49	3/49	8/49	1/49	10/49
E	5/49	5/49	0	0	10/49
$\sum_{P.j}$	15/49	15/49	10/49	9/49	1

*– حساب مصفوفة التباينات من الأعمدة r_{ik} وفق الصيغة التالية:

$$r_{ik} = \sum_{i=1}^{1} P_{ij} . P_{ik} \Big/ P_i . \sqrt{P.j.P.k}$$

مثال: حساب r_{34}:

$$r_{34} = \frac{3}{49} . \frac{4}{49} \Big/ \frac{10}{49} \sqrt{\frac{10}{49} . \frac{9}{49}} + \frac{3}{49} . \frac{4}{49} \Big/ \frac{9}{49} \sqrt{\frac{10}{49} . \frac{9}{49}} + 0 + \frac{2}{49} . \frac{1}{49} \Big/ \frac{10}{49} \sqrt{\frac{10}{49} . \frac{9}{49}} + 0$$
$$= 0.288$$

وهكذا بالنسبة لبقية أزواج المتغيرات، وترتب النتائج بالجدول التالي:

r_{ik}	EAU	SPH	MSO	MRO
EAU	0.447	0.360	0.171	0.069
SPN		0.343	0.201	0.171
MSO			0.270	0.288
MRO				0.386

ومنه نجد أن مجموع القطر يساوي $T_r = 1.446$

● **حساب القيم الخاصة – النسبة المئوية للتباينات:**

إن أول قيمة مميزة مساوية لـ/1/ الصحيح، مجموع القيم الخاصة التالية مساوية إلى $0.446 = 1 - 1.446$ ومنه نجد أن:

	القيم الخاصة	النسبة المئوية ٪	الجمع التراكمي
λ_1	1		
λ_2	0.4143	92.9%	92.9%
λ_3	0.0307	6.89%	99.78%
λ_4	0.001012	0.23%	100%

وبتقسيم كل منها على 0.446 تساوي $\frac{0.4143}{0.446} \times 100 = 0.9289$.

وبالتالي يمكن أن نكتفي بالقيمتين الخاصتين الأولى والثانية لأنهما مقربتان من /1/.

وباستخدام الصيغة التالية نجد الأشعة الخاصة:

VP_1	VP_2
0.5765	0.2884
0.2475	-0.5956
-0.3536	0.6786
-0.6947	-0.3187

*** – دراسة الأعمدة Les colonnes:**

تحديد إحداثيات المتغيرات /الأعمدة/. إحداثيات المتحول j تبعاً لمحور السينات F_α يحسب بالصيغة التالية:

$$a_j = \alpha_j \sqrt{\lambda_{k+1}/P.j}$$

حيث إن:

α_j: عبارة عن مكونات j للشعاع المميز المقابل.

λ_{k+1}: القيم الخاصة المقابلة للعامل K. وبما أن القيمة الخاصة الأولى مساوية للواحد الصحيح، إذاً نبحث عن القيمة الخاصة الثانية.

مثال:

– متحول الوسط المائي: لحساب إحداثياته على المحور الأول السينات يساوي إلى:

$axe1 = 0.5765\sqrt{0.4143 \times 49/15} = 0.671$ المحور الأول.

إحداثياته على المحور العمودي الثاني: يساوي إلى:

$axe2 = 0.2884\sqrt{0.0307 \times 49/15} = 0.091$ المحور الثاني.

- متحول وسط الأشنيات:

المحور الأول $axe1 = 0.2475\sqrt{0.4143\times 49/15} = 0.290$.

المحور الثاني $axe2 = -0.5956\sqrt{0.0307\times 49/15} = -0.189$.

وهكذا بالنسبة لبقية المتحولات كما في الجدول (5):

جدول (5): يبين إحداثيات الأوساط على المحورين الأول والثاني لـــــ AFC.

المحاور / متغيرات	المحور الأول axe1	المحور الثاني axe2
EAU	0.671	0.091
SPH	0.290	-0.189
MSO	-0.503	0.263
MRO	-1.042	-0.130

- **تفسير النتائج Interprétation:**

حساب $\cos^2$ للمتحولات بالنسبة لموقعها على المحاور:

تسمح بدراسة الاقتراب بين المتغيرات والمحاور، أي بمعزل عن الروابط بين المتغيرات والعوامل.

إن $\cos^2$ للمتحول j مع العامل h يحسب وفق الصيغة التالية:

$$\cos^2(j,h) = \frac{a_j^2(h)}{\sum_{j=1}^{P} a_j^2(P)}$$

حيث إن: P: عدد العوامل المأخوذة بالاعتبار.

$a_{j(h)}$: إحداثيات المتحول j بالنسبة للمحور axe h.

- فمثلاً بالنسبة للوسط المائي:

المحور الأول $axe1 = \frac{0.671^2}{0.671^2 + 0.091^2} = 0.982$.

المحور الثاني $axe2 = \frac{0.091^2}{0.671^2 + 0.091^2} = 0.018$.

وهكذا بالنسبة لبقية المتحولات كما في الجدول (6):

جدول (6): يبين قيم cos^2 للمتحولات على المحورين الأول والثاني لــــــ**AFC**.

Cos^2	محور axe1	محور axe2
EAU	0.082	0.018
SPH	0.702	0.298
MSO	0.785	0.215
MRO	0.985	0.015

- **مشاركة المتحولات في التباين المفسر بواسطة المحاور Les contributions:**

إن مشاركة المتحول j في التباين المفسر بوساطة المحور K يمكن تحديده بالعلاقة التالية:

$$C_{(j,k)} = 100a_j^2 \quad \text{أو} \quad 100.a_j^2.\frac{P.j}{\lambda_{k+1}}$$

مع المحور الأول axe1: $100 \times 0.5765^2 = 33.2\%$.

مع المحور الثاني axe2: $100 \times 0.2884^2 = 8.3\%$.

وهكذا بالنسبة لباقي المتحولات نجد أنها تساوي:

%	محور axe1	محور axe2
EAU	33.2	8.3
SPH	6.2	35.5
MSO	12.4	46.0
MRO	48.1	10.2

- **دراسة الأسطر في جدول البيانات Les lignes:**

حساب إحداثيات الوحدات الإحصائية/الأسطر:

إن محور السينات للوحدة الإحصائية i يقع على المحور F_k ويحسب من العلاقة التالية:

$$\beta_i = \frac{1}{P_i}.\frac{1}{\sqrt{\lambda_{k+1}}}.\sum_{j=1}^{C} P_{ij}.a_j$$

حيث إن: C: عدد الأعمدة.

λ_{k+1}: القيمة الخاصة المقابلة.

a_j: إحداثيات المتحول تبعاً للمحور F_k.

1- بالنسبة للنوع A:

إحداثيات على المحور axe1:

$$axe1 : \frac{49}{10}.\frac{1}{\sqrt{0.4143}}\left(0.671\frac{1}{49} + 0.290 \times 2/49 + (-0.503).\frac{3}{49} + (-1.042) \times 4/49\right)$$
$$= 0.688$$

إحداثيات على المحور الثاني axe2:

$$axe2:\frac{49}{10}.\frac{1}{\sqrt{0.0307}}\left(0.091\frac{1}{49}+(-0.189).2/49+0.263.\frac{3}{49}+(-0.130).4/49\right)$$
$$=-0.0101$$

وهكذا بالنسبة لبقية الإحداثيات للأنواع الأخرى كما في الجدول (7):

جدول (7): يبين إحداثيات الأنواع على المحورين الأول والثاني لـــــــ AFC.

الإحداثيات	محور axe1	محور axe2
A	-0.688	-0.010
B	-0.880	-0.065
C	0.500	0.238
D	0.234	0.112
E	0.746	-0.277

- **تفسير نتائج دراسة الأسطر (الأنواع):**

- مربعات تجيب $\cos^2$ الأنواع المدروسة (الوحدات الإحصائية بالنسبة للمحاور، أي بمعنى أنها تسمح بدراسة الروابط المتواجد بين مختلف الوحدات الإحصائية والعوامل الأخرى.

تجيب زاوية الوحدة الإحصائية i بالنسبة للمحور F_k تساوي:

$$\cos^2_{(i,k)}=\frac{\beta^2_{i(k)}}{\sum_{j=1}^{P}\beta^2_{i(P)}}$$

حيث إن P: عدد العوامل المأخوذة بالاعتبار.

$\beta_{i(k)}$: إحداثيات الملاحظة أو النوع i بالنسبة للمحور F_k.

مثال: $\cos^2$ النوع E:

بالنسبة للمحور الأول: $axe1:\frac{0.746}{0.746^2+(-0.277)^2}=0.879$ المحور الأول.

المحور الثاني: $axe2:\frac{(-0.277)}{0.746^2+(-0.277)^2}=0.121$ المحور الثاني.

وهكذا بالنسبة لباقي الأنواع، نجد أنها تساوي كما في الجدول (8):

جدول (8): يبين قيم cos^2 للأنواع على المحورين الأول والثاني لــــ **AFC**.

cos^2	محور axe1	محور axe2
A	0.999	0.000
B	0.994	0.006
C	0.815	0.185
D	0.814	0.186
E	0.079	0.121

- **مشاركة الوحدات الإحصائية في التباين المفسر بواسطة المحاور:**

إن مشاركة النوع i في التباين المفسر بواسطة المحور F_k معبرّ عنه كنسبة مئوية، ويساوي إلى :

$$C_{(i,k)} = 100.\beta_i^2.\frac{P_i}{\lambda_{k+1}}$$

– نسبة مشاركة النوع A في المحور الأول axe1:

$$axe1{:}100.(-0.688)^2.\left(\frac{10}{49}\right)^2 \Big/ 0.4143 = 23.3\%$$

$$axe2{:}100.(-0.01)^2.\left(\frac{10}{49}\right)^2 \Big/ 0.0307 = 0.1\%$$

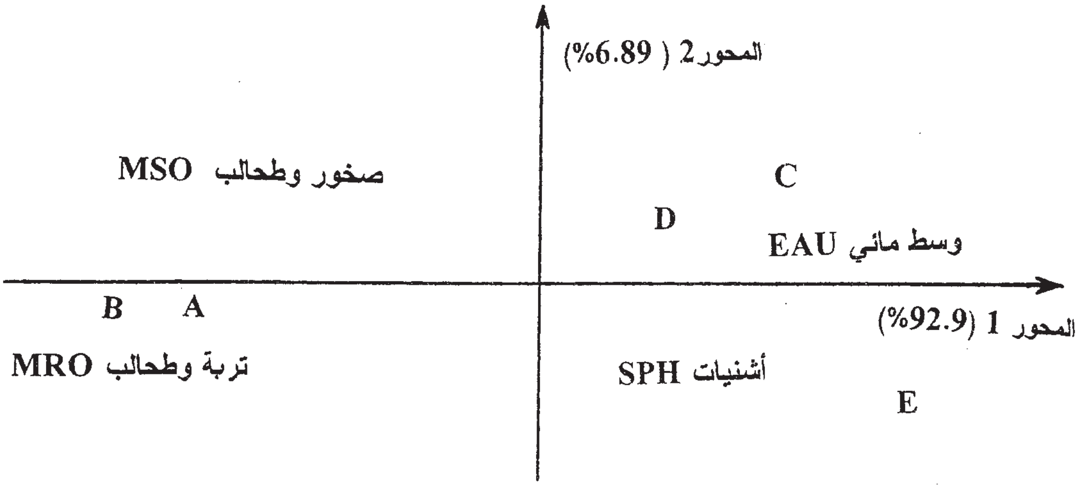
وهكذا بالنسبة لبقية الأنواع، نجد أنها تساوي كما في الجدول (9):

جدول (9): يبين قيم التباين المفسر للأنواع بواسطة المحورين الأول والثاني لــــ **AFC**.

%	محور axe1	محور axe2
A	23.3	0.1
B	34.3	2.9
C	12.3	37.7
D	2.7	8.3
E	27.4	51.1

– **التمثيل البياني المزدوج Plan factoriel 1×2/الأعمدة × والأسطر:**

استناداً إلى إحداثيات نقاط الانتشار: الأعمدة والأسطر واسقاطاتها على المخطط العاملي 2× 1 وفق الشكل (5):



الشكل (5): يبين المخطط العاملي "متحولات × أنواع" للمحورين الأول والثاني معاً لـــــــAFC.

• **النتيجة النهائية Conclusion:**

المحور I : يقابل متحول المياه EAU بالنسبة لـSPH، والمتحولين SPH وMRO، في حين المتحولان الوسط المائي EAU وMRO هما المتحولان اللذان يشتركان بدرجة كبيرة في تحديد التباين المفسر بواسطة المحاور، أي 23.2% و48.1%.

– إن قيم تجيب الزوايا القريبة من الواحد هما 0.982 و0.985 تعبر بالواقع عن روابط ذات أهمية مع العامل المدروس بواسطة هذا المحور، أما ما يتعلق بالأنواع (الأسطر)، المحور I يقابل الأنواع A وB أي الأنواع البرية بالنسبة للأنواع المائية (C,D,E)، وبالتالي المحور I يحقق تدرجاً بالنسبة للظروف الرطبة، أي المياه.

المحور II: يتصف بمشاركة فعالة للمتحولين MSO وSPH (35.5% و 46%) والأنواع C وE (37.7% و 51.1%) في هذه الأثناء القيم القليلة لـ$\cos^2$ لا تسمح بتفسير العامل المقابل.

4- دراسة مقارنة بين طريقة التحليل التقابلي AFC وطريقة ACP:

لنفترض أن لدينا جدول بيانات متعددة المتغيرات، في أغلب الأحيان لا يمكن تحليلها مباشرة، بصورة دقيقة، تعكس حقيقة وطبيعة البيانات،والظواهر التي تمثلها، كما لا يمكن الحصول منها على تمثيل بياني محدود الأبعاد، يمكن تفسيره بشكل صحيح.

إن استعمال الطرق الإحصائية الكلاسيكية لمقارنة المتغيرات بطريقة الأزواج طريقة معقدة وطويلة، إضافة لذلك، تتطلب توافر عدة شروط تطبيقية: طبيعة توزع البيانات والقوانين الاحتمالية

الخاصة لها،...الخ، بينما طرق التحليل العاملي لا تتطلب إخضاع البيانات لأية تحويلات تقربها من التوزع الطبيعي. وينصح (Benzecri, 1973) باستخدام موديل رياضي ملائم لمعلاجة البيانات وفق عدة مقاييس، بهدف الوصول إلى طريقة تستخلص البنى الهيكلية للبيانات دون حذف مخل بجوهرها.

فالتحليل العاملي بطريقة ACP وبطريقة AFC يعطي حلولاً ناجعة للمسألة المطروحة، وتظهر الطريقتان أهمية الإحصاء الوصفي في تمثيل مجموعة من المشاهدات لعدد من المتغيرات ضمن فضاء شعاعي محدود الأبعاد /المتجهات/، بمعنى أن العلاقة بين الوحدات الإحصائية أو بين المتغيرات تُعَرض في أفضل وجه ممكن.

وتشكل هذه الطرق أسلوباً رائعاً يسمح بتوضيح البنى الهيكلية للبيانات، وتوزيعها بصورة معنوية وفعالة يحددها للمتغيرات أو الأفراد التي تظهر مشاركة هامة في تشكيل المتجهات الرئيسية، وذلك بتفسير تدريجي للمتجهات، والتأكيد على العوامل المسؤولة عن هذا التوزع كما في الجدول (10).

جدول (10): مقارنة بين طريقة التحليل AFC وطريقة التحليل ACP.

1- الجداول Tableaux	
طرق التحليل العاملي التقابلي AFC	طرق التحليل بطريقة المكونات الأساسية ACP
- يطبق على جداول التوافق - المركبة التكرارية	- يطبق حصراً على جداول القياسات الكمية
2- الأصل Origine	
Kendall-Benzecri, Cordier	Spearman, Brut, Pearson, Hotelling
3- الأهداف Buts	
تمثيل البيانات متعددة المتغيرات على مسطح فضائي محدود الأبعاد. مع الحد الدنى من فقدان المعلومات المتعلقة بالمتغيرات المدروسة	
4- البيانات Données	
كمية أو نوعية	بيانات كمية فقط
5- الجداول Tableaux	
جداول القياسات - التكرار - الجداول النوعية غائب/موجود، الجداول المنطقية - التوافق	جداول القياسات
6- توزيع البيانات Distribution des données	
لا تتطلب طرق التحليل العاملي أية فروض سابقة حول توزع البيانات	
7- الخصائص والمزايا	
الخاصة (1): تعمل على تنظيم البنى الهيكلية ضمن فضاءات عاملة متعددة الاتجاهات	
الخاصة (2): تستند AFC إلى قياس كاي مربع Distance du χ^2	تستند إلى قياس اقليدس Distance euclidienne تتصف بضرورة حساب الدرجة المعيارية لتخفيف حالة عدم التجانس حيث إن: $X = \frac{x_{ij} - \bar{x}_j}{s_j}$

تابع جدول (10)

8- التمثيل البياني Représentation

AFC

تمثيل مزدوج للمتغيرات × المشاهدات على نفس المسطح (المخطط البياني)، وهذا يسهل عملية التفسير وتحليل النتائج

ACP

تعتمد على تمثيل بياني مستقل للمتغيرات عن المشاهدات أو الأفراد

9- المبدأ العام Principe

AFC

يسمح التحليل بطريقة AFC بوصف عوامل الاقتران بين الأسطر والأعمدة لجدول البيانات، مع الأخذ بيعن الاعتبار الأهمية الخاصة لكل منهما، وتوضيح ذلك بين العناصر المكونة لهذه الأسطر، أو الأعمدة على قاعدة معامل الاقتران.

نقطة الابتداء تستند إلى أساس مصفوفة جداء العوامل الاحتمالية المرجحة مع مراعاة التوزيع الحالي للأفراد. وفيما يلي التسلسل في معالجة الجداول الإحصائية AFC.

المتغيرات P عنصراً

		J	T
المشاهدات	i	n_{ij}	$n_i = \sum_j n_{ij}$
	T	$h_j = \sum_i h_{ij}$	$n = \sum_{I,J} n_{ij}$

↓

	J	T
i	$P_{ij} = {n_{ij}}/{N}$ احتمال اقتران i مع j	$P_i = \sum_j P_{ij}$ احتمال وجود i مهما كانت j
T	$P_j = \sum P_{ij}$ احتمال وجود j مهما كانت i	

↓

	K
j	r_{jk}

$$r_{jk} = \sum_{i=1}^{1} P_{ij} . P_{ik} / P_i . \sqrt{P_j . P.k}$$

يعطى التباين بين الأعمدة بالنسبة للتباين الكلي ويعطى على شكل مصفوفة قطرية.

ACP

في فضاء التغيرات بطريقة ACP. نقطة ابتداء التحليل لمصفوفة التباين المشترك Matrice covariance إذا كان F_P, F_2, F_1 تمثل عوامل مختلفة، كل عامل يحدد بواسطة موديل وفق النموذج التالي:

$$\vec{x} = \alpha_1 \vec{F_1} + \alpha_2 \vec{F_2} + ... + \alpha_P \vec{F_P}$$

مع أن $\sum_{i=1}^{P} \alpha_i^2$

والتي تشكل مكونات العوامل الخاصة (المناحي الخاصة والتي تقابلها القيم الخاصة (الخاصة) $\lambda_P ...$ λ_2, λ_1 المصفوفة معامل الارتباط مع الإشارة إلى λ_K تقابل تباين العامل K قبل أن يحدد موقع النقاط والتي تمثل ارتباطها مع المحاور العاملية وكل إحداثية من الإحداثيات نضرب بواسطة الجذر التربيعي للقيم الخاصة المقابلة لها وفق التالي:

$$a_K = \alpha_K \sqrt{\lambda_K}, a_\lambda, a_2 ... a_P$$

والتي تمثل مشاركة المتغيرات x تبعاً للمحاور العاملية، مجموع القيم الخاصة تقابل التباين الكلي لسحابة نقاط الانتشار، قبل تسلسل العوامل نحسب مشاركة كل واحد بالنسبة للتباين الكلي Inertie totale ويعطى على شكل نسبة مئوية

$$C_i = 100 {\lambda_i}/{\varepsilon_\lambda}$$

المقترحات والتوصيات:

نستخلص مما تقدم أن مختلف طرق التحليل العاملي متعدد المتغيرات تؤدي بالنتيجة إلى الملاحظات التالية:

- وصف دقيق لجداول البيانات متعددة المتحولات.
- تبويب ودراسة البنى الهيكلية الكامنة في صلب جداول البيانات.
- وصف مجموعات المشاهدات الناجمة عن التحليل العاملي.

- تعتبر هذه التقنيات من أفضل الأساليب الإحصائية التي تصور بشكل دقيق وبواسطة مخططات بيانية البنى الهيكلية للبيانات من خلال توزيع البيانات وتحديد العوامل المسؤولة عن التوزيع.
- إن إسقاط انتشار الأعمدة /المتغيرات/ والأسطر /المشاهدات/ في آن واحد على نفس المخطط العاملي يسمح لنا بدراسة دقيقة لتوزيع المشاهدات وفق اقترانها مع بعضها البعض أو وفق ارتباطها مع بعض المتغيرات. من هنا نبين أهمية هذه التقنيات في الدراسات الإحصائية ولاسيما أن توفر الحاسوب كأداة لتسهل هذا الأسلوب وعلى وجه الخصوص في الدراسات البيئية أو الاقتصادية أو الاجتماعية والتي تعالج التأثير المتبادل من عدة متغيرات في آن واحد وعلى مجموعة أو عدة مجموعات من المشاهدات.

REFERENCES المراجع

- ANDERSON T. W., 1958. – An Introduction to Multivriate Analysis. Wiley. New – York.
- BARTLETT M., 1933. – On the Theory of Statistical Regression. Proc. R. soc. Edinbu;, 53.
- BENZECRI J. P. et al. 1973. – L'analyse des données. Tom 1. La taxonomie. Tom 2. l'analyse des correspondances. Dunod, Paris, Fr. 615 et 619 PP.
- BENZECRI J. P. et al. 1980. – Pratique de l'analyse des données: analyse des correspondences. Exposées élémentaire. Bordas, Paris. 424P.
- CAILLIEZ F., PAGES J. P. 1976. – Introduction à l'analyse des données. SMASH. PARIS. 616 P.
- ESCOFIER B. PAGES j. P. 1988. – Analyses factorielles simples et multiples. Dunod. Paris. 242P.
- HOTELLING H. 1936. Relation between tow sets of variates. Biometrika, 28, 129-149.
- JAMBU M. 1973. Introduction à l'analyse des données: les methods de classification automatique. Consmation, 3.
- LEGENDRE L. et LEGENDRE P. 1984; - Ecologie numérique 1: Le traitement multiple des données écologiques. Masson, Paris et les presses de l'Université LAVAL. Québec 2 éme éd.
- PEARSON K. 1901. – On lines and planes of calsest fit to systems of poits in spaces. Phill. Mag. 2.
- PIELOU E. C. 1975. – Mathematical Ecology. Wiley intersciences, New – York.
- RAO C. A., 1952. – Advanced Statistical methods in biometric research. Wiley J. New – York.
- SPEARMAN C. 1904. – General intelligence objectively determined and measured. Am J. Psychol. 15.