الفصل الرابع : التحليل العاملي للتوفيقات "Analyse factorielle des correspondances AFC"

تعتبر طريقة التحليل العاملي للتوفيقات " AFC" من أهم الطرق الاستكشافية في التحليل المتعدد الابعاد وأكثرها شيوعا في والدراسات الابحاث العلمية. فعلا غرار طريقة تحليل المركبات الاساسية والتي تستعمل في حالة البيانات الكمية فان هذه الطريقة تعتبر الانسب أو تمثل الحل الامثل في حالة البيانات الكيفية لذلك فهي لا تقل أهمية عنها. في هذا الفصل سوف نتناول طريقة التحليل العاملي للتوفيقات في حالة متغيرين او البسيط ثم التحليل العاملي للتوفيقات المتعدد.

IV- 1 تقديم طريقة التحليل العاملي للتوفيقات " AFC"

هي طريقة احصائية استكشافية متعدد الابعاد تسمح بتحويل جدول من الارقام الى تمثيل رسومي (خريطة) mapping و تستعمل هذه الطريقة كثيرا في حالة الاستبيانات.

كما انها تهتم في العموم بجداول الاقتران المعرفة بمتغيرتين كيفيتين، حيث تقوم بجمع الى عدد اقل من الابعاد لمجموع المعلومات الاولية من خلال التركيز ليس على القيم المطلقة ولكن على القيم النسبية بين المتغيرات. هذا التخفيض مفيد جدا خصوصا عندما يكون عدد الابعاد الاولية مرتفع.

ان مفهوم التخفيض مشترك بين جميع تقنيات التحليل العاملي أي حيث نستخلص العوامل لذلك فطريقة AFC تقدم الخصوصية على عكس ACP في توفير مساحة تمثيل مشتركة للمتغيرات والافراد معا . لذلك تعتبر AFC حقيقة طريقة تحليل للمركبات الاساسية ACP خاصة ، تنجز على مجموع التكرارات النسبية الافقية والعمودية من اجل الحصول على خريطة مشتركة (سحابة نقاط) مما يسهل امكانية دراسة التوافق بين مختلف مستويات تصنيف المتغيرين (Modalités) وبالتالي دراسة العلاقة بين هذين المتغيرين ، لذلك فهي امتداد ل ACP من حيث ملائمتها لمعالجة المعطيات الكيفية بدون قيود قوية.

الهدف الرئيسي AFC إذا هو قراءة للمعطيات المحتواة داخل مساحة (فضاء) متعددة الابعاد عن طريق تخفيض لأبعاد هذا الفضاء مع الحفاظ بأكبر قدر ممكن من المعلومات المحتواة في المساحة الابتدائية.

1-1-IV جدول الاقتران (التوافق) "Tableau de contingence"

- **المعطيات (Donnée)**

تطبق طريقة تحليل العاملي للتوفيقات على جداول الاقتران أو التوافق. وهو عبارة عن طريقة معينة، خاصة للعرض الاني لمتغيرين وصفيين سواء كانا اسميين او ترتبين .لذلك فهو جدول ذو مدخلين تتوزع فيه مجموع مفردات مجتمع ما وفقا لخاصيتين معينتين y و x .حيث تتموضع احداهما على السطر والاخرى على العمود ، مما يمكن من اقتران مستويات تصنيفها "Modalités" وفقا لمجموعة من القيم تعرف بالتكرارات والتي دائما ما تكون اعداد موجبة طبيعية.

- **الترميز (Notations)**

ليكن لدينا المتغيرتين الوصفيتين x و y بحيث ان مستويات تصنيف كل متغير هي على ترتيب : $a_1, a_2, \dots, a_i \dots, a_n$ و $b_1, b_2, \dots, b_j, \dots, b_p$. بحيث تتم مشاهدتها في ان واحد من اجل n مفردة (لها نفس الوزن $\frac{1}{n}$) . وباعتبار ان المتغيرة x تتموضع على الاسطر والمتغيرة y تتموضع على العمود فان تقاطع السطر والعمود يعطي لنا جدول التكرارات التالي:

	b_1	b_2	$\dots$	b_j	$\dots$	$\dots$	b_p	
a_1	n_{11}	n_{12}	$\dots$	n_{1j}	$\cdots$	$\cdots$	n_{1p}	$n_{1.}$
a_2	n_{21}	n_{22}	$\cdots$	n_{2j}	$\cdots$	$\cdots$	n_{2p}	$n_{2.}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$			$\vdots$	$\vdots$
a_i	n_{i1}	n_{i2}	$\cdots$	n_{ij}	$\cdots$	$\cdots$	n_{ip}	$n_{i.}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$			$\vdots$	$\vdots$
$\vdots$	$\vdots$	$\vdots$		$\vdots$			$\vdots$	$\vdots$
a_q	n_{q2}	n_{q2}	$\cdots$	n_{qj}	$\cdots$	$\cdots$	n_{qP}	$n_{q.}$
	$n_{.1}$	$n_{.2}$	$\dots$	$n_{.j}$	$\dots$	$\dots$	$n_{.P}$	$n_{..}$

$Z_{(q,p)} =$

حيث:

$\boldsymbol{n_{ij}}$: هي التكرارات الموافقة ل a_i b_j والذي يمثل عدد المفردات التي من اجلها x تأخذ القيمة a_i و y القيمة b_j .

$\boldsymbol{n_{i.}}$: التكرار الهامشي الموافق للسطر i ويعبر عن عدد المفردات التي من اجلها x تأخذ قيمة i والتي تعطي بالعلاقة التالية: $n_{i.} = \sum_{j=1}^{p} n_{ij}$

$\boldsymbol{n_{.j}}$: التكرار الهامشي الموافق للعمود j ويمثل عدد المفردات التي من اجلها y تأخذ القيمةj والتي تعطي بالعلاقة التالية : $n_{.j} = \sum_{i=1}^{q} n_{ij}$

كل سطر وكل عمود يوافق عينة فرعية خاصة من مجموع المفردات $n_{..}$ أي : $n_{..} = \sum_{i=1}^{q} \sum_{j=1}^{p} n_{ij}$

مثال:

جدول الاقتران التالي يبين تقاطع التغيرين الوصفيين " الشعبة المدروسة في الجامعة و فئة المهنيين الاجتماعيين" حيث تشمل الاولى اربعة توجهات " مستويات تصنيف" وفي القانون, العلوم, الطب والتكوين المهني والثانية وتشمل خمس قطاعات مهنية وهي مزارع , رئيس, اطار سامي , موظف و عامل .

	Droit	Sciences	Médecine	IUT	Totale
Exp.agri	80,000	99,000	65,000	58,000	302
Patron	168,000	137,000	208,000	62,000	575
Cadre.sup	470,000	400,000	876,000	79,000	1825
Employé	145,000	133,000	135,000	54,000	467
Ouvrier	166,000	193,000	127,000	129,000	615
Totale	1029	962	1411	382	3784

فعلى سبيل المثال:

نسبة ابناء الاطارات السامية من مجموع الافراد ككل = 1825/3784 = 48.2 %

نسبة الاطفال الذين اختاروا شعبة الطب من مجموع الافراد ككل = 3784/1411 = 37.3 %

نسبة الافراد الذين اختاروا مسار الطب وهم أبناء لإطارات سامية في الدولة = 876/1411 = 23.2 (فكرة المصاحبة ، الارتباط).

نسبة الافراد الذين اختاروا مسار الطب مع العلم انهم من أبناء لإطارات سامية في الدولة = 876/1825 = 48.0% (فكرة السببية).

تهتم طريقة تحليل العاملي للتوفيقات خصوصا بالتكرارات النسبية وليس المطلقة .لذلك فان المجاميع الهامشية على السطر والعمود جد مهمة وبفرض وجود ارتباط (علاقة) بين المتغيرتين y و x ، فان وصف هذه العلاقة وشرحها يفرض علينا تحويل الجدول الابتدائي من جدول التكرارات المطلقة الى جدول التكرارات النسبية .

- **جدول الاحتمالات (التكرارات النسبية):**

بقسمة جميع خانات الجدول الابتدائي على مجموع المفردات فانه يمكن الحصول على الجدول التالي:

$f_{(q,p)} =$

	b_1	b_2	...	b_j	...	...	b_p	
a_1	f_{11}	f_{12}	...	f_{1j}	$\cdots$	$\cdots$	f_{1p}	$f_{1.}$
a_2	f_{21}	f_{22}	$\cdots$	f_{2j}	$\cdots$	$\cdots$	f_{2p}	$f_{2.}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$			$\vdots$	$\vdots$
a_i	f_{i1}	f_{i2}	$\cdots$	f_{ij}	$\cdots$	$\cdots$	f_{ip}	$f_{i.}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$			$\vdots$	$\vdots$
$\vdots$	$\vdots$	$\vdots$		$\vdots$			$\vdots$	$\vdots$
a_q	f_{q2}	f_{q2}	$\cdots$	f_{qj}	$\cdots$	$\cdots$	f_{qP}	$f_{q.}$
	$f_{.1}$	$f_{.2}$	...	$f_{.j}$	...	...	$f_{.P}$	$f_{..}$

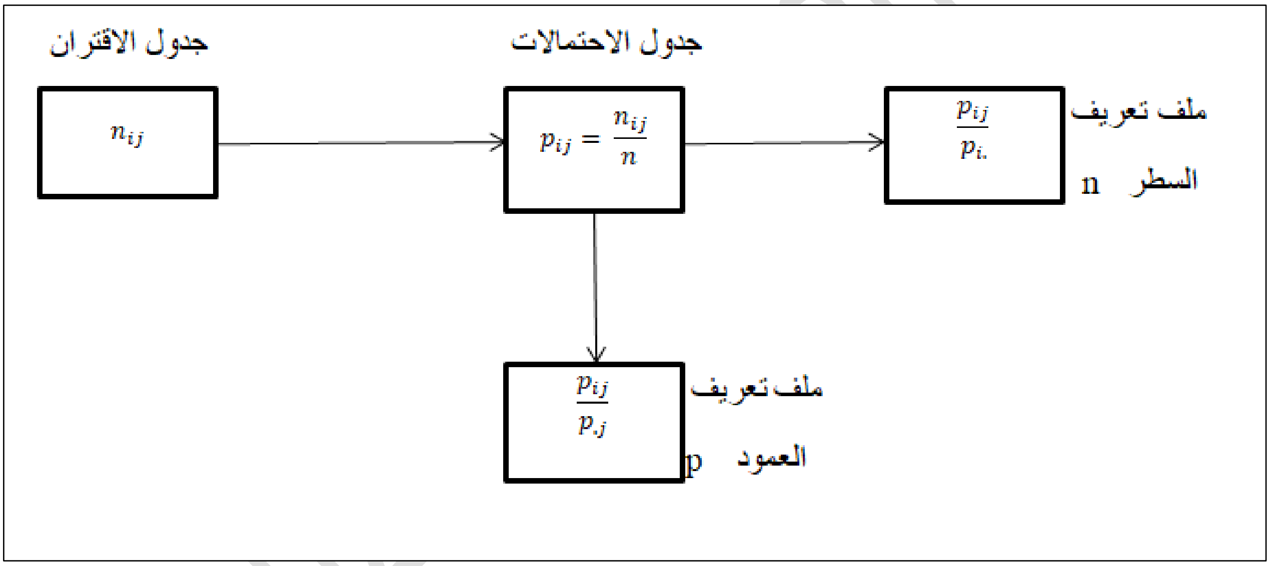
حيث :

$\boldsymbol{f_{ij}}$: هي التكرارات النسبية الموافقة ل ai و b_j / $f_{ij} = \frac{n_{ij}}{n} = p_{ij}$

$\boldsymbol{f_{i.}}$: التكرار الهامشي النسبي الموافق للسطر / $f_{i.} = \frac{\sum_{j=1}^{p} n_{ij}}{n} = p_{i.}$

$\boldsymbol{f_{.j}}$: التكرار الهامشي النسبي الموافق للعمود / $f_{.j} = \frac{\sum_{i=1}^{q} n_{ij}}{n} = p_{.j}$

التحول من الجدول الى جدول التكرارات $Z_{(q,p)}$ الى جدول الاحتمالات $f_{(q,p)}$ يمثل اول خطوة من خطوات AFC . حيث من خلاله يمكن الحصول على جدولين اخرين بمعنيين مختلفين : جدول بروفيل ن الاسطر جدول التكرارات النسبية للسطر وجدول بر وفيلات العمود حيث يمكن تمثيلها على النحو التالي :



في جدول الاقتران ، لا تحمل الكلمات المفردات والمتغيرات نفس الدلالة كالجداول المستطيلة لل " ACP" .في الواقع ، في جداول الاقتران ، تمثل الخطوط والاعمدة مجموع مستويات التصنيف للمتغيرتين x و y على الترتيب. وذلك للحفاظ على التجانس في التمثيل بالنسبة للطريقتين (AFC وACP)، سوف يفترض بالاتفاق ان مجموع مستويات التصنيف n للمتغير x على السطر تحمل اسم المفردات وان p مستوى تصنيف للمتغيرة y على العمود تحمل اسم المتغير.

وعليه فإننا نلاحظ ايضا ان الجدول الابتدائي (Z) يمكن كتابته بدون مبالاة مع مستوى التصنيف x على السطر . او على العمود نفس الشيء بالنسبة ل y بدون ان تتغير طبيعة الجدول ومع ذلك ففي هذه الحالة ، فان جداول الملفات (البروفايل= profils) على الصف والعمود لم تعد لها نفس المعنى .

طريقة AFC هي نفسها ACP بالقياس ، فانطلاقا من المصفوفة Z او من تحويلاتها الى مصفوفات بروفيل ، يمكن تحليل المعلومة المحتواة في ذلك الجدول من خلال فضاءين :

2-1-IV ملامح السطر والمسافة بين البروفايلات '' Profils lignes et la distance entre profils ''

ويقصد به الفضاء R^p "للمتغيرات" ويمثل مستويات تصنيف الاعمدة اين يمكن ان نمثل سحابة النقاط q للمفردات « Modalités ligne » كل مفردة من اجل الاحداثيات $\frac{p_{ij}}{p_{i.}}$ في هذا الفضاء نستعمل جدول بروفيل السطر(Tableau des profils ligne). في الفضاء R^p،سوف نهتم بالقيم النسبية للنقاط الفردية ، أي ملامح الخطوط وبالتالي نكون امام هذه المصفوفة :

يمكن حساب المسافة بين نقطتين من مصفوفة تشكيلات الاسطر x_{ij} (جدول الاسطر) مثلا تشكيلة السطر i والتي نرمز لها بالرمز l_i :

$$\left(\frac{n_{i1}}{n_{i.}} \quad \frac{n_{i2}}{n_{i.}}, \dots, \frac{n_{ip}}{n_{i.}}\right) \Leftrightarrow \left(\frac{f_{i1}}{f_{i.}} \quad \frac{f_{i2}}{f_{i.}}, \dots, \frac{f_{ip}}{f_{i.}}\right) / \quad \frac{n_{ij}}{n_{i.}} = \frac{f_{ij}}{f_{i.}}$$

Tableaux des profils lignes =

	b_1	b_2	…	b_j	… …	b_p		
a_1	$\frac{p_{11}}{p_{1.}}$	$\frac{p_{12}}{p_{1.}}$	…	$\frac{p_{1j}}{p_{1.}}$	… …	$\frac{p_{1p}}{p_{1.}}$	1	
a_2	$\frac{p_{21}}{p_{2.}}$	$\frac{p_{22}}{p_{2.}}$	…	$\frac{p_{2j}}{p_{2.}}$	… …	$\frac{p_{2p}}{p_{2.}}$	1	
⋮	⋮	⋮		⋮		⋮	⋮	
$\boldsymbol{a_i}$	$\boldsymbol{\frac{p_{i1}}{p_{i.}}}$		…	$\boldsymbol{\frac{p_{ij}}{p_{i.}}}$	… …	$\boldsymbol{\frac{p_{ip}}{p_{i.}}}$	**1**	**احداثيات النقطة i في الفضاء R^p**
⋮	⋮	⋮		⋮		⋮	⋮	
⋮	⋮	⋮		⋮		⋮	⋮	
a_q	$\frac{p_{q1}}{p_{q.}}$	$\frac{p_{q2}}{p_{q.}}$	…	$\frac{p_{qj}}{p_{q.}}$	… …	$\frac{p_{qp}}{p_{q.}}$	1	
	$\frac{p_{ij}}{p_{i.}}$	$\frac{p_{ij}}{p_{i.}}$	…	$\frac{p_{ij}}{p_{i.}}$	… …	$\frac{p_{ij}}{p_{i.}}$	q	

حيث أن:

$$P(X=i \,/\, Y=j) = \frac{n_{ij}}{n_{i.}}$$

المسافة المقدرة بين بروفايلات السطر مثنى مثنى هي الفروق بين مستويات تصنيف الاسطر تعطى عن طريق كاي تربيع والتي تعطى بالعلاقة التالية:

$$d^2{}_\chi(i,\grave{i}) = \sum_{j=1}^{p} \frac{n_{..}}{n_{.j}} \left(\frac{n_{ij}}{n_{i.}} - \frac{n_{\grave{i}j}}{n_{\grave{i}.}}\right)^2$$

مثال:

Profils (lignes) جدول ملامح الخطوط					
	Droit	**Sciences**	**Médecine**	**IUT**	**Somme**
Exp.agri	0,265	0,328	0,215	0,192	1
Patron	0,292	0,238	0,362	0,108	1
Cadre.sup	0,258	0,219	0,480	0,043	1
Employé	0,310	0,285	0,289	0,116	1
Ouvrier	0,270	0,314	0,207	0,210	1
Moyenne	0,279	0,277	0,311	0,134	1

من خلال الجدول يمكن استنتاج مايلي:

- نسبة الاطفال الذين اختاروا شعبة العلوم من مجموع الافراد ككل = 962/3784 = 4.25%
- نسبة منهم أبناء لإطارات السامية في الدولة مع العلم انهم اختاروا مسار العلوم = 400/1825 = 21.9%
- نسبة منهم أبناء العمال مع العلم انهم اختاروا مسار العلوم = 615/193 = 31,4 %

السؤال: على العموم فان ابناء الاطارات السامية وابناء العمال قاموا باختيار نفس المسار التعليمي؟ ما هو الحال بالنسبة لأبناء اصحاب العمل (الرؤساء) والعمال ؟ للإجابة على هذا السؤال فانه يكفي حساب المسافة بين البروفايلات:

$$d^2(cadre, ouvrier) = \frac{1}{0,272}(0,258-0,270)^2 + \frac{1}{0,254}(0,219-0,314)^2 + \frac{1}{0,373}(0,480-0,207)^2 + \frac{1}{0,101}(0,043-0,210)^2 = 0,5109$$

$$d^2(cadre, patron) = \frac{1}{0,272}(0,258-0,292)^2 + \frac{1}{0,254}(0,219-0,238)^2 + \frac{1}{0,373}(0,480-0,362)^2 + \frac{1}{0,101}(0,043-0,108)^2 = 0,084611$$

إن اختيار المسارات لأبناء أصحاب العمل (الرؤساء) أقرب (مشابه) لتلك لأبناء الإطارات السامية. مما هم من أبناء العمال.

بالنسبة للملامح الهامشية (Moyenne) = وتمثل الاختيارات للمسار الدراسي دون الاخذ بعين الاعتبار أو التمييز بين القطاعات المهنية (CSP) = البروفايلات المتوسطية أو ما يعرف بالمتوسط التقديري لملامح الأسطر مثال على ذلك:

$$0.272=\frac{1}{3784}\times(302\times0.256+575\times0.292+1825\times0.258+467\times031+615\times0.270)$$

- **المسافة الى الاصل (المركز) أو المسافة الى الملامح الوسطية "Distance à l'origine "**

بالاعتماد على العلاقة السابقة يمكن حساب المسافة الى الاصل :

$$d^2(cadre)=\frac{1}{0{,}272}(0{,}258-0{,}272)^2+\frac{1}{0{,}254}(0{,}219-0{,}254)^2+\frac{1}{0{,}373}(0{,}480-0{,}373)^2+\frac{1}{0{,}101}(0{,}043-0{,}101)^2=0{,}0693$$

Poids, distances et distances quadratiques à l'origine, inerties et inerties relatives (lignes)					
	Poids (relatif)	Distance	Distance²	Inertie	Inertie relative
Exp.agri	0,080	0,413	0,170	0,01360	0,161
Patron	0,152	0,058	0,003	0,00050	0,006
Cadre.sup	0,482	0,263	0,069	0,03342	0,395
Employé	0,123	0,174	0,030	0,00372	0,044
Ouvrier	0,163	0,453	0,205	0,03340	0,395

كتحليل عام ومن وجهة نظر التموضع بالنسبة للمتوسط فانه يمكن القول أن نسبة ابناء العمال وكذا ابناء المزارعين تمثل تشكيلات الاختيارات (لملامح) الاكثر "اختلافا" من بين جميع الاطفال.

كتحليل معمق فانه ومن وجهة نظر التموضع (المسافة) مثنى مثنى فانه يمكن القول أن ابناء العمال وكذا ابناء الفلاحين تمثل تشكيلات (هياكل) الاختيار الاكثر قربا أي الاكثر تشابها. ومن ناحية أخرى فان أبناء الاطارات السامية تبتعد كثيرا (أي تختلف كثيرا) عن ابناء القطاعين السابقين.

تعطي طريقة "AFC" تمثيلا بيانيا للتموضع الذي يرجع الى مستويات التصنيف وهذا جيد جدا من الناحية العملية في حالة كون عددها كبير جدا.

- **العطالة لملامح (بروفايلات) الاسطر " Distance à l'origine et inertie "**

يمكن تقدير العطالة او ما يعرف بالتباين وهذا بروفايلات الاسطر عن طريق العلاقة التالية :

العطالة (Inertie) = المسافة الى الأصل(« Distance à l'origine « DISTO) × وزن مستوى التصنيف (Poids de la modalité)

مثال : العطالة (cadre) = وزن $(\text{cadre}) \times d^2(\text{cadre}) = \frac{1825}{3784} \times 0{,}069 = 0{,}0395$

وعليه تكون العطالة بالنسبة للقطاعات الاخرى على النحو التالي :

	Poids (relatif)	**Distance²**	**Inertie**	**Inertie relative**
Exp.agri	0,080	0,170	0,01360	0,161
Patron	0,152	0,003	0,00050	0,006
Cadre.sup	0,482	0,069	0,03342	0,395
Employé	0,123	0,030	0,00372	0,044
Ouvrier	0,163	0,205	0,03340	0,395
			Totale	0,085

مجموع العطالة او العطالة الكلية = مقدار المعلومة الموجودة في أجمالي البيانات والذي يعتير مؤشرا بالغ الاهمية .

3-1-IV ملامح العمود "Profils colonnes" والمسافة بين البر وفيلات:

ويقصد به الفضاء R^q ل "المفردات" اين يمكن تمثيل P متغيرة . كل متغيرة لها الاحداثيات $\frac{p_{ij}}{p_{.j}}$. في هذا الفضاء سوف نستعمل جدول البر وفيلات العمود . لهذا سنكون أمام هذه المصفوفة :

	b_1	b_2	…	$\boldsymbol{b_j}$	… …	b_p	
a_1	$\frac{p_{11}}{p_{.1}}$	$\frac{p_{12}}{p_{.2}}$	…	$\frac{p_{1j}}{p_{.j}}$	… …	$\frac{p_{1p}}{p_{.p}}$	$\frac{p_{ij}}{p_{1.}}$
a_2	$\frac{p_{21}}{p_{.1}}$	$\frac{p_{22}}{p_{.2}}$	…	$\frac{p_{2j}}{p_{.j}}$	… …	$\frac{p_{2p}}{p_{.p}}$	$\frac{p_{ij}}{p_{2.}}$
⋮	⋮	⋮		⋮		⋮	⋮
a_i	$\boldsymbol{\frac{p_{i1}}{p_{.1}}}$		…	$\boldsymbol{\frac{p_{ij}}{p_{.j}}}$	… …	$\boldsymbol{\frac{p_{ip}}{p_{.p}}}$	$\frac{p_{ij}}{p_{i.}}$
⋮ ⋮	⋮ ⋮	⋮ ⋮		⋮ ⋮		⋮ ⋮	⋮ ⋮
a_q	$\frac{p_{q1}}{p_{.1}}$	$\frac{p_{q2}}{p_{.2}}$	…	$\frac{p_{qj}}{p_{.j}}$	… …	$\frac{p_{qp}}{p_{.p}}$	$\frac{p_{ij}}{p_{q.}}$
	1	1	…	**1**	… …	1	Q
				احداثيات j في الفضاء			

R^q

Tableau des profils colonnes

=

أي جدول الاعمدة او تشكيلات الاعمدة Y_{ij} ، مثل : تشكيلة العمود j والتي يرمز لها بالرمز c_j :

$$\frac{n_{ij}}{n_{i.}} = \frac{f_{ij}}{f_{i.}} \quad / \quad \begin{pmatrix} \frac{f_{1j}}{f_{.j}} \\ \frac{f_{2j}}{f_{.j}} \\ \vdots \\ \frac{f_{qj}}{f_{.j}} \end{pmatrix} \Leftrightarrow \begin{pmatrix} \frac{n_{1j}}{n_{.j}} \\ \frac{n_{2j}}{n_{.j}} \\ \vdots \\ \frac{n_{qj}}{n_{.j}} \end{pmatrix}$$

حيث أنه من خلا ل الجدول يمكن تحديد:

$$P\,(y = j \,/\, x = i) = \frac{n_{ij}}{n_{.j}}$$

المسافة المقدرة بين بروفايلات العمود مثنى مثنى هي الفروق بين مستويات تصنيف الاعمدة و تعطى عن طريق كاي تربيع والتي تعطى بالعلاقة التالية:

$$d_{\chi}^{2}\,(j, \grave{j}) = \sum_{i=1}^{q} \frac{n}{n_{i.}} \, \left(\frac{n_{ij}}{n_{.j}} - \frac{n_{i\grave{j}}}{n_{.\grave{j}}}\right)^2$$

مثال:

Profils (colonnes) بروفايلات العمود					
	Droit	**Sciences**	**Médecine**	**IUT**	**Moyenne**
Exp.agri	0,078	0,103	0,046	0,152	0,095
Patron	0,163	0,142	0,147	0,162	0,154
Cadre.sup	0,457	0,416	0,621	0,207	0,425
Employé	0,141	0,138	0,096	0,141	0,129
Ouvrier	0,161	0,201	0,090	0,338	0,197
Somme	1	1	1	1	1

من خلال الجدول يمكن استنتاج مايلي:

- نسبة ابناء الاطارات السامية من مجموع الافراد ككل = 42,5 %
- نسبة أبناء لإطارات السامية في الدولة مع العلم انهم اختاروا مسار القانون = 45,7 %
- نسبة أبناء لإطارات السامية في الدولة مع العلم انهم اختاروا مسار التكوين المهني = 20,7 %

السؤال: بالنسبة للقطاعات ماهي التي تملك تشكيلات متشابهة؟

للإجابة على هذا السؤال يمكن تطبيق المسافة بين :

$$d^2(droit.Sciences) = \frac{1}{0.080}(0{,}078 - 0{,}103)^2 + \frac{1}{0{,}152}(0{,}163 - 0{,}142)^2 + \cdots \ldots = 0{,}024$$

$$d^2(droit, médecine) = \frac{1}{0.080}(0{,}078 - 0{,}046)^2 + \frac{1}{0{,}152}(0{,}163 - 0{,}147)^2 + \cdots \ldots = 0{,}118$$

نلاحظ أن الملامح السيسيولوجيا للطلبة في القانون هي الاكثر قربا الى طلبة العلوم منه الى طلبة الطب.

- **المسافة الى الاصل (المركز) أو المسافة الى الملامح الوسطية "Distance à l'origine "**

بالاعتماد على العلاقة السابقة يمكن حساب المسافة الى الاصل :

Poids, distances et distances quadratiques à l'origine, inerties et inerties relatives (colonnes)					
	Poids (relatif)	**Distance**	**Distance²**	**Inertie**	**Inertie relative**
Droit	0,272	0,069	0,005	0,001	0,015
Sciences	0,254	0,165	0,027	0,007	0,082
Médecine	0,373	0,305	0,093	0,035	0,409
IUT	0,101	0,644	**0,414**	0,042	0,494

Totale	0,0846

يمثل البروفايل أو الملامح الاجتماعية لطلبة التكوين المهني الاكثر اختلافا من بين جميع طلبة التخصصات الاخرى هذا بالنظر الى المسافة الى الاصل (DISTO).

أما بالنسبة للعطالة فتمثل المسار "IUT" و مسار "Médecine" مستويات التصنيف الاكثر تعبيرا للمعلومة.

أما في ما يتعلق بالعطالة الكلية فنلاحظ انها متماثلة مع تلك المتعلقة بدراسة ملامح السطر وهذا انه من الطبيعي اننا نقوم بتحليل نفس الجدول ولكن من وجهة نظر مختلفة.

4-1-IV تقدير العلاقة بين المتغيرات الكيفية (ارتباط الخطوط – الاعمدة واختبار الاستقلالية ل كاي تربيع) :

إن الهدف من طريقة تحليل العاملي للتوفيقات هو تحليل العلاقة الموجودة بين متغيرتين كيفيتين- في حالة وجود اكثر من متغيرتين نوعيتين فإننا نلجأ الى تحليل العاملي للتوفيقات المتعددة AFCM والتي تعتبر كتعميم فقط لهذه الطريقة - لذا ، قبل تنفيذ او تطبيق AFC يجب التأكد اولا من وجود هذه العلاقة لهذا ، هناك رسوم بيانية (اعمدة بيانية للبر وفيلات) وكذا خصائص رقمية (مؤشر مربع الكيدو ومشتقاته) .والتي تسمح بإبراز مثل هذه الرابطة عند وجودها فبالاستناد الى اختبار مربع كاي للاستقلالية (H_0 المتغيرين X و Y مستقلان وهذا يستلزم أن الاحتمال المشترك P(Y= j et X=i)= P(Y=j) P(X=i)))، يمكن اختبار ما اذا كان هناك ارتباط كبير من عدمه بين هذين المتغيرتين النوعيتين.

تجدر الاشارة انه وبالقياس مع طريقة ACP وفان هذه العلاقة تعطى عن طريقة المسافة الاقليدية بين نقاط سحابات النقاط للفضائيين R^p و R^q. وعلى العكس من ذلك فانه بالنسبة AFC ، لا نستعمل هذه المسافة الاقليدية .بتعبير ادق نستخدمها ولكن بعد اجراء تحويل قبلي لا حداثيات نقاط السحابات في الفضائيين :

حيث يعطى بالعلاقة التالية:

$$\chi^2_{obs} = \sum_{i=1}^{q}\sum_{j=1}^{p}\frac{(effectif\ observé - effectif\ théorique)^2}{effectif\ théorique}$$
$$= \sum_{i}^{q}\sum_{j}^{p}\frac{(n_{ij}-e_{ij})^2}{e_{ij}} = \sum_{i}^{q}\sum_{j}^{p}\frac{\left(n_{ij}-\left(\frac{n_{i.}n_{.j}}{n_{..}}\right)\right)^2}{\frac{n_{i.}n_{.j}}{n_{..}}}$$
$$= \sum_{i}^{q}\sum_{j}^{p} n_{..}\ \frac{(probabilité\ observé - probabilité\ théorique)^2}{probabilité\ théorique} = n\,\phi^2$$

حيث ان :

χ^2_{obs}: وتمثل قيمة كاي تربيع المحسوبة او الملاحظة .

effectif observé: وتمثل التكرارات المشاهدة (n_{ij}) .

effectif théorique: وتمثل التكرارات النظرية (e_{ij}) والتي تحسب انطلاقا من التكرارات الهامشية وفقا للعلاقة التالية :

$$e_{ij} = \frac{n_{i.}n_{.j}}{n}$$

ϕ^2 : معامل فأي مربع يمثل شدة العلاقة من خلال الفارق بين الاحتمالات النظرية والمشاهدة. حيث كلما اقتربت المسافة الى 0 هذا يعني ان هناك استقلالية .كلما كانت عظمية أي هناك ترابط بين المتغيرتين.

كما تعطى ايضا المسافة عن طريق الكيدو باعتبار التكرارات النسبية بالعلاقة التالية:

$$D(\chi^2_{obs}) = \sum_i^q \sum_j^p \frac{(nf_{ij} - nf_{i.}f_{.j})^2}{nf_{i.}f_{.j}} = \sum_i^q \sum_j^p n \frac{(f_{ij} - f_{i.}f_{.j})^2}{f_{i.}f_{.j}}$$

وعليه فمن خلال هذه العلاقة يمكن معرفة مساهمة كل خانة من خانات جدول الاقتران في قيمة كاي تربيع المحصلة من خلال العلاقة التالية :

$$/C_{qp} = \frac{{r_{qp}}^2}{\chi^2} = \frac{(n_{qp} - e_{qp})^2}{e_{qp}} = \frac{(f_{ij} - f_{i.}f_{.j})^2}{f_{i.}f_{.j}}$$

$r_{ij} = \frac{n_{ij} - e_{ij}}{\sqrt{e_{ij}}}$ $\sim$ approximativement N(0,1) C.-à-d. Au risque 5%, $|r_{ij}| > 2 \rightarrow$ significatif.

حيث أن:

r_{qp} هي الاخطاء المعيارية

C_{qp} هي المساهمة النسبية في قيمة كاي تربيع (أو المساهمة في المعلومة التي تحملها كل خانة من خانات الجدول).

تتم مقارنة قيمة كاي تربيع المحسوبة (المشاهدة) مع قيمة كاي تربيع النظرية (الجدولية) في جدول كاي تربيع وهذا من اجل درجة حرية « dll » المساوية ل (p-1) (q-1) وكذا مستوى دلالةα والذي عادة ما تؤخذ القيمة 5 % حيث :

- اذا كان : $\chi^2_O > \chi^2_T$ فإننا نقبل الفرضية البديلة وهذا ما يؤكد على وجود الارتباط ما بين المتغيرين الوصفيين .
- اذا كان : $\chi^2_O < \chi^2_T$ فإننا نقبل الفرضية الصفرية أي أن هناك استقلالية بين المتغيرين .

مثال: فيما يلي سوف نعرض عليكم مجموع النتائج المحصل عليها والتابعة للمثال السابق والتي تتعلق بحسابات اختبار كاي تربيع :

Effectifs théoriques التكرارات النظرية

	Droit	Sciences	Médecine	IUT	Total
Exp.agri	82,124	76,777	112,612	30,487	302,000
Patron	156,362	146,181	214,409	58,047	575,000
Cadre.sup	496,280	463,967	680,517	184,236	1825,000
Employé	126,993	118,725	174,138	47,144	467,000
Ouvrier	167,240	156,350	229,325	62,085	615,000
Total	1029	962	1411	382	3784

يتم حساب القيم النظرية من خلال عبارة الكيدو السابقة ,كمثال يتم الحصول على القيمة 82,1 عن طريق :

$$82{,}1 = \frac{302 \times 1029}{3784}$$

Test d'indépendance entre les lignes et les colonnes (Khi²)	
Khi² (Valeur observée)	320,266
Khi² (Valeur critique)	21,026
DDL	12
p-value	< 0,0001
Alpha	0,05

من خلال الجدول يتضح أن من أجل $\propto = 5\ \%$ فإننا نرفض الفرضية الصفرية التي تدل على استقلالية المتغيرين X و Y. اي ان هناك ارتباط بين المتغيرين.

ملاحظة : من أجل تأكيد النتائج المتوصل اليها عن طريق اختبار كي تربيع فانه يجب مايلي:

- يجب ان تكون على الاقل 80 % من التكرارات النظرية ($e_{ij} \geq 5$) لأنه وبكل وضوح فان وجود قيم ضعيفة ل e_{ij} الى تضخيم قيمة الكيدو.
- يجب ان تكون قيمة n كبيرة مما يعني أن قيمة كاي دائما ما تكون ذات دلالة احصائية (لان درجة الحرية لا تأخذ بعين الاعتبار قيمة n) .
- يجب ان تكون قيمة كاي تربيع : $0 \leq \chi^2 \leq n \times \min(q-1,\ p-1)$
- تعتبر النتائج التي يعطيها معامل كاي تربيع منقوصة من حيث الدلالة لذلك غالبا ما نلجأ لتدعيمها عن طريق اختبارات اخرى والتي تتمثل في : اختبار (*Tschuprow*) و اختبار (*V de Cramer*) والتي تأكد اعتدالية قانون كاي تربيع. ويعطي كل منهما بالعلاقة التالية :

$$\text{Test de Tschuprow : } t = \sqrt{\frac{\chi^2}{n\sqrt{(q-1)(p-1)}}}$$

$$\text{Test de V de Cramer : } v = \sqrt{\frac{\chi^2}{n\min(q-1,p-1)}}$$

كلا الاختبارين يتغيران ما بين 0 و ال 1 كما يقومان على تعديل اختبار كاي بالاعتماد على درجة الحرية.

من خلال قيمة كاي تربيع فانه يمكن اشتقاق قيمة اخرى والتي تتمثل في معامل فاي تربيع والذي يعبر عن العطالة الكلية حيث :

$$\text{PHI}-2 \quad \phi^2 = \frac{\chi^2}{n} = \frac{320.3}{3784} = 0.0846$$

مثال: من خلال المثال السابق فان قيمة كل من t و v هي على الترتيب: 0, 156 و 0, 168 حيث أن :

$$v = \sqrt{\frac{\chi^2}{3784\ \min(5-1,\ 4-1)}} = 0,168,\ t = \sqrt{\frac{320,3}{3784\sqrt{(5-1)(4-1)}}} = 0,156$$

Coefficients d'association	
Coefficient	**Valeur**
Phi de Pearson	0,291
Coefficient de contingence	0,279
V de Cramer	0,168
T de Tschuprow	0,156
Tau de Goodman et Kruskal (L/C)	0,033
Tau de Goodman et Kruskal (C/L)	0,027

- **المساهمة النسبية في قيمة كاي تربيع:**

من خلال المساهمة النسبية في قيمة كاي تربيع يظهر لنا أن مسار التكوين المهني هو المسار الذي يساهم أكثر في اجمالي المعلومة بنسبة % 22,52 .

Khi² par case المساهمة النسبية في قيمة كاي تربيع					
	Droit	Sciences	Médecine	IUT	Total
Exp.agri	0,02	2,01	6,29	7,75	16,06
Patron	0,27	0,18	0,06	0,08	0,59
Cadre.sup	0,43	2,75	17,53	18,77	39,49
Employé	0,80	0,54	2,75	0,31	4,39
Ouvrier	0,00	2,68	14,26	22,52	39,46
Total	1,52	8,16	40,88	49,44	

Résidus (Pearson) الأخطاء المعيارية				
	Droit	Sciences	Médecine	IUT
Exp.agri	-0,234	2,536	-4,487	4,983
Patron	0,931	-0,759	-0,438	0,519
Cadre.sup	-1,180	-2,970	7,494	-7,753
Employé	1,598	1,310	-2,966	0,998
Ouvrier	-0,096	2,931	-6,757	8,492

الاشارة الموجبة (+) فتعني الجاذبية أما الاشارة السالبة (-) فتعني النفور.

حيث يمكن الحصول على جدول التنافر والتجاذب عن طريق حساب مؤشر التجاذب والتنافر (indice d'attraction et répulsion) عن طريق العلاقة التالية:

$$i_{qp} = \frac{p\ (x = q\ \&\ y = p)}{p(x = q) \times p(y = p)} \ \# \ \frac{o_{qp}}{e_{qp}}$$

حيث أن اذا كان $i_{qp} < 1$ فهذا يعني ان هناك تنافر أما اذا كان $i_{qp} > 1$ فهذا يعني ان هناك تجاذب .

5-1-IV تقنية عمل الطريقة AFC وتحديد التنافر والتجاذب " Technique de la méthode et la détermination de l'attraction-répulsion"

بعد حساب قيمة كاي تربيع للتأكيد العلاقة بين الاسطر والاعمدة تجدر الاشارة الى ان طريقة AFC يمكن ان تعطي معلومات مهمة حتى وان كان χ^2 غير معنوي فانه سوف تنتقل الى عمل تحليل للمركبات الاساسية "ACP" لكل من تشكيلات السطر والعمود كل على حدا.

حيث تقوم بتطبيق طريقة ACP على الجدول الخاص بتشكيلات "ملامح =Profiles" الاسطر (المفردات) وفي هذه الحالة هي الاسطر الخاصة بجدول الاقتران بمعنى اخر مستويات تصنيف المتغيرة x ثم نقوم بالتمثيل البياني للمفردات ، أي مستويات تصنيف x (في هذه الطريقة لتحليل للمركبات الاساسية الخاصة "ACP particulier" ، لن نهتم بتمثيل المتغيرات) وعليه نحصل على تمثيل واحد فقط هذا اذا احتفظنا ببعدين فقط ، عدة تمثيلات بيانية على خلاف ذلك.

من ناحية اخرى ، فإننا نقوم بعمل ACP للجدول الخاص بتشكيلات العمود (Profils-colonnes) المفردات في هذه الحالة هي عبارة عن اعمدة جدول الاقتران بمعنى اخر هي مستويات التصنيف للمتغيرة y ، ثم نقوم بالتمثيل البياني لمجموع المفردات ، أي تمثيل مستويات التصنيف للمتغيرة y .

تظهر ان هذين الطريقتين لتحليل المركبات الاساسية الخاصيتين متوافقتين وهذا الامر طبيعي ، لأنه يتم استخراج بياناتهم من نفس جدول الاقتران وبالتالي فانه من الشرعي القيام بتراكب superposer التمثيليتين البيانيين . وبالتالي نحصل على رسم بياني من نوع سحابة النقاط nuage de points (او مجموعة من الرسوم البيانية اذا احتفظنا بأكثر من بعدين) يمثل في نفس الوقت مستويات التصنيف للمتغيرة x و y .

ان تفسير هذا الرسم البياني ، الذي من اجله نتوفر على مجموعة من المؤشرات، التي تسمح بشرح العلاقة بين المتغيرتين المدروسين .على وجه الخصوص ، سوف نشرع في دراسة التوقيفات بين مستويات التصنيف ل x و y لذلك سوف نتطرق الى مقياس نقاط الاسطر والاعمدة ,مركز الثقل لسحابة الاسطر والاعمدة والمحاور الاساسية.

بالاعتماد على المثال السابق يمكن شرح العلاقة بين هذين المتغيرين الكيفيين وهذا بطرح مجموعة من الاسئلة:

- من هم (ابناء القطاعات) CSP الذين يتخذون نفس المسار التعليمي (الشعبة) ؟

- ما هي المسارات التعليمية (الشعب) التي لديها هياكل (أبناء القطاعات المختلفة) CSP متقاربة؟
- أي قطاع CSP ينجذب إلى أي مسار (شعبة) ؟ أو العكس أي ما هو المسار الذي له جاذبية أكثر الى

لأي من القطاعات CSP؟

نود الحصول على نضرة شاملة ومختصرة. مثالية هي احد التمثيلات البيانية.

- **تحليل البروفايلات أو الملامح :**

وتتمثل في البحث عن أول عامل يجعل من الممكن وضع مستويات التصنيف للمتغيرتين (Xو Y) بالكيفية الاكثر تشتتا بحيث تكون في أغلب الاحيان متناثرة (بحيث يمكن تمييزها عن بعضها البعض على افضل وجه ممكن).

- **بالنسبة لملامح الاسطر** :

يتم رسم مختلف نقاط مستويات تصنيف الاسطر عن طريق ايجاد احداثيات كل نقطة (أربع نقاط في المثال السابق والتي تمثل مستويات التصنيف) على ذلك العامل وهذا عن طريق تعظيم قيمة (λ) والتي تمثل تباين نقاط مستويات والتي تعطى بالعلاقة التالية :

$$\lambda_1 = \sum_{i=1}^{q} \frac{n_{i.}}{n} \times F_{i1}^2$$

حيث:

F_{i1} هي الاحداثيات العاملية على مستوى العامل الاول لنقاط مستويات تصنيف الاسطر.

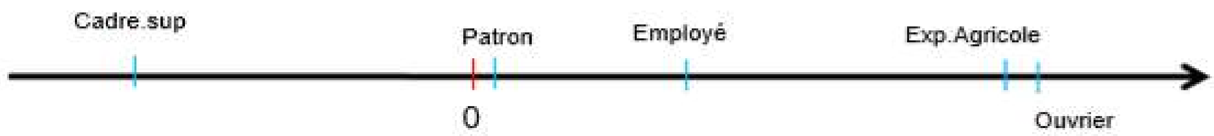
$\frac{n_{i1}}{n}$ هو وزن كل مستوى تصنيف

التي تشير الى مصداقية العامل (الاخلاص للواقع) $\frac{\lambda_1}{\phi^2}$. يمكن استنتاج القيمة λ_1بالاعتماد على قيمة

مثال:

يمكن الحصول على إحداثيات مستويات تصنيف الصف من خلال جميع نقاط الأعمدة (والعكس صحيح): المتوسط المرجح مفرغًا (منزوعا منه) من جذر القصور الذاتي.

Coordonnées principales (lignes) احداثيات نقاط مستويات تصنيف الاسطر				
	Poids (relatif)	F1	F2	F3
Exp.agri	0,080	0,410	0,026	-0,038
Patron	0,152	0,020	-0,027	0,047
Cadre.sup	0,482	-0,263	0,016	-0,006
Employé	0,123	0,142	-0,097	-0,021
Ouvrier	0,163	0,451	0,040	0,009



- **بالنسبة لملامح العمود :**

نفس الشيء بالنسبة لنقاط تصنيفات العمود (والتي تتمثل في اربع نقاط في المثال) حيث يمكن حساب قيمة (λ) بالعلاقة التالية :

$$\lambda_1 = \sum_{j=1}^{p} \frac{n_{.j}}{n_{..}} \times G_{j1}^2$$

حيث:

G_{j1} هي الاحداثيات العاملية على مستوى العامل الاول لنقاط مستويات تصنيف الاعمدة .

نشير هنا الى ان (λ) المحسوبة هي من نفس القيمة في حالة ملامح السطر.

مثال:

Coordonnées principales (colonnes) احداثيات نقاط مستويات تصنيف الاعمدة				
	Poids (relatif)	F1	F2	F3
Droit	0,272	0,028	-0,061	0,017
Sciences	0,254	0,160	-0,003	-0,038
Médecine	0,373	-0,303	0,030	0,005
IUT	0,101	0,640	0,061	0,031

مثال:

$\Phi^2 = \frac{\chi^2}{n} = 0,0846$ والتي تمثل قيمة العطالة الكلية.

Inertie totale	0,085		
Valeurs propres et pourcentages d'inertie			
	F1	F2	F3
Valeur propre	0,082	0,002	0,001
Inertie (%)	97,350	2,013	0,638
% cumulé	97,350	99,362	100,000

من خلال الجدول يتضح أن (0.0824/0.0846) = 97.4% من اجمالي المعلومة المتاحة والتي تم استرجاعها بواسطة العامل (المحور الاول).

ملاحظة: 100٪ من المعلومات = الجدول الأولي للمسافات بين الملفات الشخصية (البروفايلات). في الواقع، يقدم الموضع العاملي صورة مخلصة إلى حد ما هنا. طريقة تحليل المعاملات للتوفيقات تصبح فعالة (حاسمًة) عندما يكون عدد مستويات التصنيف مرتفعًا.

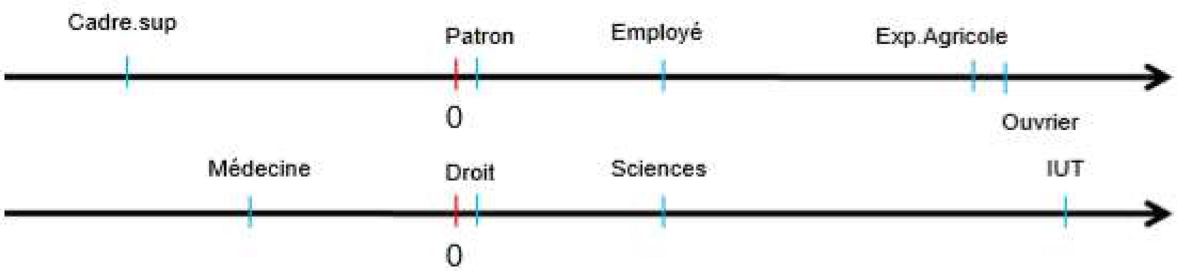
القيمة التي اظهرتها λ_1 هي التي سمحت برسم التمثيل البياني المحوري لمختلف النقاط سواء للأسطر والاعمدة (représentation axiale) لان نسبة المعلومة التي تبقت هي 3 % استطاعت ان تفسرمنها المركبة الثانية (المحور الثاني) حوالي 2 % فقط ($\lambda_2 = 0.002$) وهذا ما يعتبر بالشيء القليل جدا مما أدى الى اهماله.

- قيمة λ_1 هي نفسها المحسوبة سواء بالنسبة لمستويات تصنيف السطر أو العمود ;
- موافق لتحديد المواقع النسبية لمستويات تصنيف الصفوف (أعمدة على التوالي) ;

- تحليل الارتباطات " Analyse des associations"

ماذا يمكن أن يقال من حيث ارتباطات (التجمعات) الصفوف و الأعمدة؟ هل تقاربهم على المحور هو مؤشر؟

من خلال التمثيلين البيانين السابقين (التمثيل المحوري) وبإسقاط أحدهما على الاخر يمكن القول :



- هل أبناء العمال ينجذبون الى مسار التكوين المهني؟
- هل أبناء أرباب العمل ينجذبون الى مسار القانون؟

مثال :

Profils (lignes) ouvrier جدول ملامح الخط للعمال				
	Droit	**Sciences**	**Médecine**	**IUT**
Ouvrier	0,270	0,314	0,207	0,210

Coordonnées principales (colonnes) sur le 1er axe احداثيات نقاط مستويات تصنيف الاعمدة بالنسبة للمحور الاول				
	Droit	**Sciences**	**Médecine**	**IUT**
Coord 1	0,028	0,160	-0,303	0,640

$$F_{i1} = \frac{1}{\sqrt{\lambda_1}} \sum_{j=1}^{p} \frac{n_{.j}}{n_{..}} \times G_{j1}$$

احداثيات "العمال" من العلاقة:

$$F_{51} = \frac{1}{\sqrt{0.0824}} (0.2699 \times 0.02799 + \cdots + 0.2098 \times 0.64017) = 0.45148$$

حوصلة:

1. التوفيق بين إحداثيات مستويات الصف / العمود قانوني بفضل تلك العلاقة
2. لكن يمكن أن يتم ذلك على العموم فقط وهذا يعني يجب تحديد وضع أحدى مستويات الخط بواسطة مقارنة بجميع طرائق العمود (والعكس صحيح).

بالعودة الى السؤالين اللذين تم طرحهما سابقا :

- هل ينجذب أطفال العمال إلىIUT ؟

الجواب: نعم .ينجذب أطفال العمال إلى IUT أكثر من جميع الطلاب.

- هل ينجذب أبناء ارباب العمل (الرؤساء) إلى القانون؟

الجواب: لا .بالمقارنة مع جميع الطلاب، فانه لا ينجذب أبناء الرؤساء الى أي قطاع على وجه الخصوص (كما هو الحال بالنسبة للموظف في مكان آخر)

كنتيجة فان تفسير الارتباطات او بعبارة اخرى الاقتران بين مختلف مستويات تصنيف السطر والعمود يستند اساسا الى مؤشر التجاذب والتنافر (Indice d'attraction- répulsion). والذي يعطى بالعلاقة التالية:

$$\widehat{\iota_{\jmath}} = \sum_{h=1}^{H} \frac{F_{jh} G_{ih}}{\sqrt{\lambda_h}}$$

هذا المؤشر يحسب انطلاقا من الجدول الاقتران الاولي.

مؤشر التجاذب والتنافر (Indice d'attraction- répulsion)				
	Droit	Sciences	Médecine	IUT
Exp.agri	0.974	1.289	0.577	**1.902**
Patron	1.074	0.937	0.970	1.068
Cadre.sup	0.947	0.862	**1.287**	0.429

Employé	1.142	1.120	0.775	1.145
Ouvrier	0.993	1.234	0.554	2.078

من خلال الجدول نلاحظ ان :

أ 2 من مستويات التصنيف تتجاذبان (تتدافعان) إذا كانت إحداثياتها لها نفس العلامة (الإشارة المعاكسة) على المحاور. حيث تظهر القيم المكتوبة بالخط العريض المستويات الاكثر تجاذبا من بين الجميع.

كون الخط هو الأكثر تمييزًا كما هي القيم عالية (بالقيمة المطلقة).

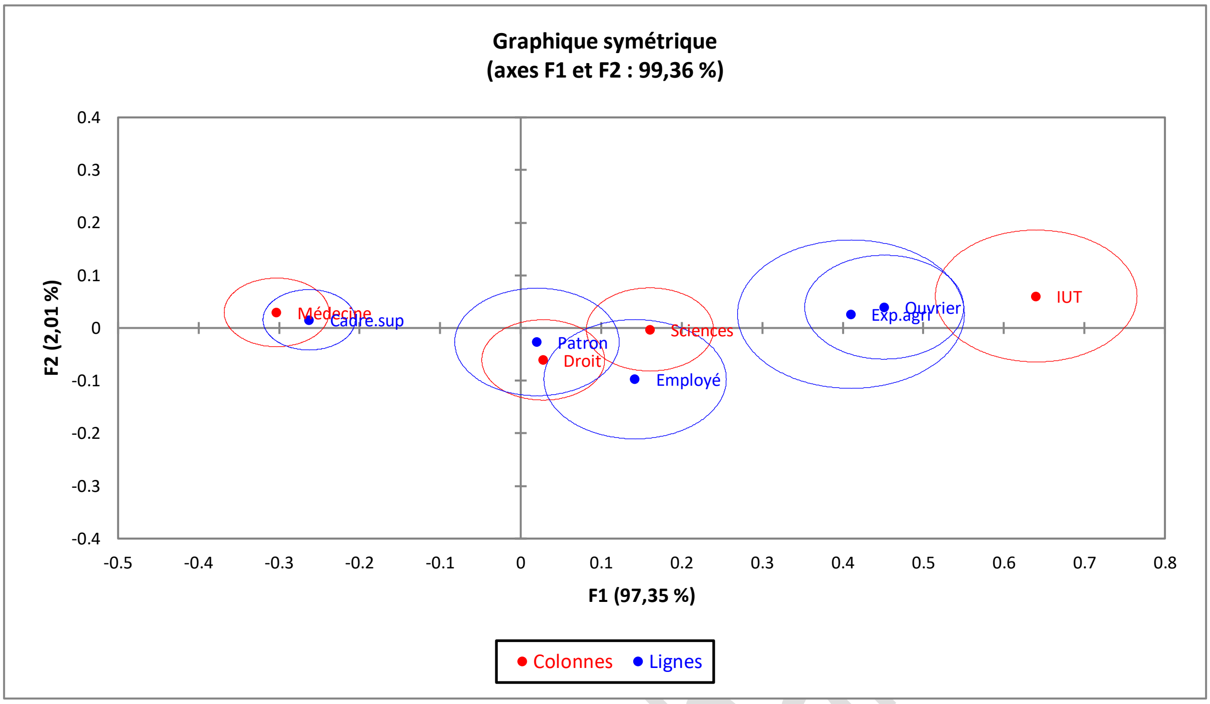
7-1-IV التمثيل الاني لمستويات تصنيف الخط والعمود في المخطط "Représentation simultanée dans le plan"

يمكن تلخيص معظم النتائج السابقة في الجدول التالي والذي يسمح بدوره بتفسير المخطط في الاسفل والذي يعبر عن سحابات النقاط الخاصة بمستويات تصنيف الاسطر والاعمدة :

Ligne	Colonne	Observée	attendue	Résidus.std	Contribution	Cumul (%)
Ouvrier	IUT	129	62.1	8.49	22.52	22.52
Cadre.sup	IUT	79	184.2	-7.75	18.77	41.29
Cadre.sup	Médecine	876	680.5	7.49	17.53	58.82
Ouvrier	Médecine	127	229.3	-6.76	14.26	73.08
Exp.agri	IUT	58	30.5	4.98	7.75	80.83
Exp.agri	Médecine	65	112.6	-4.49	6.29	87.12

يحتوي الجدول المساهمات فقط التي تكون أكبر من المتوسط بمعنى اخر المساهمات % / 100

$(q \times p) <$



من خلال الرسم التخطيطي يمكن التمييز بين المجموعات بعضها عن بعض حيث يتضح ان :

أبناء العمال وكذا المزارعين ينجذبون الى المسار التكوين المهني مقارنة بالطلاب الاخرين (ابناء اصحاب القطاعات الاخرى) كما أنهم بالمقابل ينفرون من مسار الطب . وبهذا تكون هذه النتيجة بالتناظر مع أبناء الاطارات السامية الذين ينجذبون الى الطب وينفرون من التكوين المهني, هذا التناظر يكون بالنسبة للمحور الاول . أما بالنسبة لأبناء أرباب العمل فهم لا ينجذبون نحو أي مسار مقارنة بالطلبة الاخرين غير ان ابناء الموظفين فينجذبون ولو بنسبة قليلة جدا الى المسار العلمي.

- **اختيار المحاور الاساسية " Choix des axes factoriels "**

عند اجراء طريقة تحليل العاملي بالتوفيقات (نفس الحال بالنسبة لطريقة تحليل المركبات الاساسية) من خلال جداول الاقتران والذي يحوي q صف و p عمود. فعلا سبيل المثال اذا كان q>p فان ابعاد الفضاء الذي حصل على مستواه مجموع النتائج هو p-1 ، اما اذا كان q<p فان هذا البعد هو q-1 . بصورة عامة فان مجال تعريف هذا البعد هو (p-1، q-1).

باختصار فان العدد (H) الاقصى للمحاور الذي يمكن الاحتفاظ به هو : $H_{max} = \min(q-1; p-1)$

بالاعتماد على قاعدة (Kaiser) فإننا نحتفظ دائما بالمحور الذي يوفر نسبة عطالة أعلى من ($\frac{1}{H_{max}}$). في المثال السابق نجد ان :

الحد الاقصى للمحاور هو 3 (1-4) وهذا يعني : أننا سوف نحتفظ بالمحور الذي يوفر على الاقل نسبة $33\ \% = (1/3)$. ولذلك قمنا بالاحتفاظ بالمحور الاول وجوبا أما الثاني فلانه على الاقل وجود محورين من أجل الحصول على الرسم التخطيطي.

- **الاحداثيات العاملية وجودة التمثيل :**

من أجل حساب احداثيات العاملية لمختلف مستويات التصنيف لكل متغيرة فانه نقوم بتطبيق العلاقة التالية:

$$CTR_h(p) = \frac{\frac{n_{i.}}{n}\ d_i^2\ (i)}{\sum_{i=1}^{q} \frac{n_i}{n}\ d_h^2(i)} = \frac{\frac{n_i}{n} d_h^2\ (i)}{\lambda_i}$$

- **جودة التمثيل "Qualité de représentation"**

ويقصد بها نسبة المعلومة المتعلقة بمستويات التصنيف و المفسرة من أجمالي المعلومة عن طريق المحاور الرئيسية كما هو الحال بالنسبة لطريقة تحليل المركبات الاساسية . حيث أنها العطالة (القصور الذاتي) لمستوى التصنيف على المحور مقسومًا على العطالة الكلية (الجمود الكلي) لمستوى التصنيف. وتعطى عن طريق العلاقة التالية :

$$COS_h^2(i) = \frac{\frac{n_{i.}}{n}\ d_h^2\ (i)}{\frac{n_{i.}}{n}\ d^2(i)} = \frac{d_h^2\ (i)}{d^2\ (i)}$$

حيث أن:

$$Somme_h[COS_h^2(i)] = 100\ \%$$

مثال: من خلال الجدولين الخاصين بمربع جب التمام لمستويات التصنيف نلاحظ ان جميعها ممثلة تمثيلا جيدا على مستوى المحور الاول (جميع القيم قريبة الى الواحد) عدا مستوى التصنيف للسطر والخاص ب فئة ارباب العمل (Patron) وكذا مستوى التصنيف للعمود الخاص بالمسار القانون (Droit) واللذان كان تمثيلهما سيئا على مستوى المحور الاول وهذا بقربهما من المركز (قيمة جب التمام تقريبا منعدمة) لذلك لا يمكن القول عنهما اي شيء في خصوص الارتباك (الاقتران). هذه النتيجة تعكس نسبة المعلومة التي جاءت محملة عن طريق المحور الاول (97%) . أما بالنسبة للتمثيل على مستوى المحمر الثاني فنلاحظ أن جميع مستويات التصنيف جاء تمثيلها سيئا وهذا ما يدل على نسبة المعلومة التي استطاع ان يفسرها المحور الثاني (2%).

Cosinus carrés (lignes) :			
	F1	F2	F3
Exp.agri	0,987	0,004	0,009
Patron	0,123	0,213	0,664
Cadre.sup	0,996	0,004	0,001

Employé	0,670	0,315	0,015
Ouvrier	0,992	0,008	0,000

Cosinus carrés (colonnes) :			
	F1	F2	F3
Droit	0,165	0,777	0,058
Sciences	0,948	0,000	0,052
Médecine	0,990	0,009	0,000
IUT	0,989	0,009	0,002

- **التمثيل البياني تشكيلات السطر والعمود :**
- **سحابة تشكيلات السطر :** باعتبار تشكيلات الأسطر هي مجموع المفردات فان q تشكيلة سطر تشكل سحابة نقاط من q نقطة في فضاء R^P .

من اجل كل تشكيل سطر يوافق الوزن المساوي الى التكرار الهامشي $f_{i.}$ حيث أن : $f_{i.} = \frac{n_{.j}}{n_{..}}$

نرمز ب N(i) سحابة النقاط المشكلة من تشكيلات السطر المرجحة $(l_i, f_{i.})$ حيث أن: l_i هي تشكيلة السطر i , $f_{i.}$ هو الاوزان .

- **مركز الثقل :** g_l يعطي بالعلاقة التالية: $g_l = \sum_{i=1}^{q} f_{i.} l_i$

الاحداثيات i ل g_l هي $f_{.j}$ وعليه : $\sum_{i=1}^{q} f_{i.}(\frac{f_{ij}}{f_{i.}}) = \sum_{i=1}^{q} f_{ij} = f_{.j}$ حيث أن:
g_l : التشكيل الهامشي للمتغيرة X (لها q مستوى التصنيف)

وعليه يكون شعاع مركز الثقل لسحابة الاسطر هو : $g_x = \begin{pmatrix} g_1 \\ g_2 \\ \vdots \\ g_p \end{pmatrix} = \begin{pmatrix} \frac{n_{.1}}{n_{..}} \\ \frac{n_{.2}}{n_{..}} \\ \vdots \\ \frac{n_{.p}}{n_{..}} \end{pmatrix}$

- **سحابة تشكيلات الاعمدة:**

يرمز لها بالرمز N(J) والتي تكون مشكلة من النقاط P تشكيلات الاعمدة المرجحة : $(C_j, f_{.j})$ حيث ان : مركز الثقل g_c هو التشكيل الهاشمي للمتغيرة Y التي لها P مستوى التصنيف .

وعليه يكون شعاع مركز الثقل لسحابة الاعمدة هو : $g_c = f_y = \begin{pmatrix} \frac{n_{1.}}{n_{..}} \\ \frac{n_{2.}}{n_{..}} \\ \vdots \\ \frac{n_{q.}}{n_{..}} \end{pmatrix}$

- **مفهوم العطالة في طريقة AFC :**

لقد راينا الدور الذي لعبه مفهوم العطالة او التشتت في طريقة ACP (انظر المبحث الثاني) .لذلك لا يزال هذا الدور بنفس الاهمية في طريقة AFC باعتبارها من طرق التحليل العاملي ولكن بشكل خاص. حيث انه يرتبط ارتباطا كبيرا بمؤشر كاي تربيع الذي حل محل المسافة الاقليدية بالنسبة لطريقة ACP ، وعليه فانه تعطي عطالة كل من سحابة النقاط $N_I(X)$ باعتبار الاسطر محسوبة بالنسبة لمركز الثقل f_j بالعلاقة التالية:

$$N_I(X) = \sum_{i=1}^{q} f_{i.}\, d^2_{\chi^2}\left(l_i, f_j\right) = \frac{\chi^2}{n_{..}}$$

ملاحظة: نحصل على نفس النتيجة من اجل عطالة سحابة النقاط N(Y) .

IV-1-8 بعض التعليقات على طريقة التحليل العاملي للتوفيقات:

- يهدف AFC إلى تحليل الجداول المتقاطعة الكبيرة
- تعتبر طريقة AFC وصفية
- يطبق طريقة التحليل العاملي للتوفيقات خارج جداول الاقتران ايضا
- يجب أن يكون جدولًا متقاطعًا للقيم الموجبة
- وأن مفاهيم الهوامش والملامح قابلة للتطبيق
- تهتم طريقة AFC بهياكل علاقات الصف / العمود
- يمكن أن يوفر نتائج حتى لو لم يكن χ^2 غير مهم

IV-2 التحليل العاملي للتوفيقات المتعدد ACM (التوفيقات المضاعف) " Analyse des correspondances multiples "

تم تخصيص هذا الفصل لطريقة التحليل العاملي للتوفيقات (AFC) والتي تتمثل في طريقة من طرق الإحصاء الوصفي متعدد الأبعاد الذي يجعل من الممكن تحليل الارتباط بين متغيرين نوعيين (في الاغلب فئويين). بقدر ما لا تستطيع أن تأخذ في الاعتبار متغيرين فقط ، فإن AFC محدود بشكل طبيعي (يطلق عليه أحيانًا تحليل التوفيقات الثنائية، أو حتى تحليل التوفيقات البسيط).

في الممارسة العملية ، ولا سيما في مجال التعامل مع الاستطلاعات (أو الاستبيانات) ، من النادر أن نقتصر على متغيرين (أو سؤالين). المشكلة الإحصائية التي تطرح في هذا النوع من البيانات هو تحليل الارتباط الذي قد يوجد بين عدد هائل من المتغيرات النوعية. تحليل التوفيقات المتعددة (ACM) هو الأسلوب العاملي للإحصاء الوصفي متعدد الأبعاد الذي يسمح بالتعامل مع هذه المشكلة.

من حيث المبدأ، فإن (ACM) هو (AFC) خاص. ما هي التغييرات هي ببساطة جدول البيانات التي يتم تطبيق الطريقة عليه. المشكلة الأساسية هي في الواقع معرفة ما هو الجدول الإحصائي، الذي يتقاطع على مستواه عدد هائل من المتغيرات النوعية، وبالتالي يمكن أن يعمم طاولة الاقتران. في الواقع ، فان الإجابة بالفعل هي جدول بيرت (Table de Burt). وبالتالي، فإن A.C.M هو AFC يتم تنفيذه على طاولة Burt بالنسبة إلى ثلاثة على الأقل للمتغيرات النوعية.

وبالتالي فإن كيفية تفسير نتائج طريقة A.C.Mستكون مماثلة لطريقة التفسير تلك الخاصة بـ AFC. لسوء الحظ ، بعض مؤشرات المساعدة على التفسير المستخدمة في AFC لم تعد صالحة في سياق ACM. بالإضافة إلى وجود عدد أكبر من المتغيرات يجعل التفسير أكثر دقة. لذلك يتطلب التمكن الجيد من ACM ممارسة كبيرة لهذه الطريقة (أكثر من معرفة رياضية واسعة النطاق).كجزء من هذا الفصل، سيقتصر طموحنا على تقديم الطريقة بصورة اسرع وكذا شرح آلية التفسير على مثال حقيقي بسيط نسبيًا.

IV-2-1 تذكير بخصوص طاولة بيرت

في هذه الفقرة، سوف نتناول المفاهيم التي سبق تقديمها في السابق من الفصل.

- **البيانات التي يتم العمل عليها**

البيانات التي يتم توجيهها لبناء جدول بيرت هي بالضبط تلك التي تم أخذها في الاعتبار في تحليل التوفيقات المتعددة (ACM). لذلك يجب أن يكون عددًا عشوائيًا من المتغيرات النوعية (ليكن: p, p $\geq$ 3)، مشاهدة خلال مجموعة من n فرد (العينة قيد النظر) ، لكل منهم نفس الوزن $\frac{1}{n}$. المتغيرات سيتم الإشارة إليها بواسطة $X^1, X^2, \dots, X^p$ سيتم الإشارة إلى عدد مستويات تصنيف X^j بواسطة C_j ($j = 1, \dots, p$) ، وسنقوم بفرض ان $C = \sum_{j=1}^{p} c_j$ (العدد الإجمالي لمستويات التصنيف التي تم النظر فيها ، جميع المتغيرات مجتمعة).

ملاحظة: كما في (AFC)، يمكن استخدام المتغيرات الفئوية في (ACM) متغيرات نوعية، مع مستويات تصنيف مرتبة أو غير مرتبة، أو متغيرات كمية سواء كانت منفصلة أو مستمرة). وهنا يمكن التحدث عن الفئات لتعيين مستويات التصنيف أو القيم أو الفئات. مهما كانت هذه التركيبة الفئوية

(تركيبة عددية أو ترتيبية) فإنها لن تكون مأخوذة بعين الاعتبار خلال التحليل. هذا يجعل استخدام ACM مرنًا للغاية لأنه طريقة قادر على التعامل مع أي نوع من المتغيرات

- **تعريف جدول بيرت**

نقدم هنا تعريف جدول بيرت (يسهل فهمه من خلال المثال المعطى أدناه). نذكر أن جدول بيرت هو تعميم خاص للجدول الاقتران لعدد عشوائي p من المتغيرات النوعية. جدول بيرت هو في الواقع مصفوفة مربعة (مصفوفة مربعة) c × c ، تتكون من p^2 مصفوفات فرعية. كل من المصفوفات الفرعية القطرية p ترجع إلى أحد المتغيرات p، ويشتمل على الأعداد الهامشية على القطر من X^j. المصفوفة الفرعية التي تظهر في كتلة المؤشرات (j,$\hat{j}$) حيث j ≠$\hat{j}$ هي جدول الاقتران الذي تم إنشاؤه بوضع X^j في الصفوف و $X^{\hat{j}}$ في الأعمدة. لذلك فإن جدول بيرت هو جدول متماثل.

- **التوضيح**

دعونا نأخذ احد الامثلة كما هو مضح في الجدول ادناه حيث قمنا بدراسة عينة من 797 طالب من جامعة Paul Sabatier (Toulouse III) الذين تحصلوا على DEUG A أو DEUG B (الدرجات العلمية الجامعية للطور الاول ، في سنتين) ، وهذه الدرجة فقط كانت خلال الفترة 1971-1983. تم أخذ ثلاثة متغيرات في الاعتبار: سلسلة البكالوريا (Série de Bac) والتي لها مستويين من التصنيف (C, D) ؛سن الحصول على البكالوريا (l'âge d'obtention du bac) من اجل اربع مستويات تصنيف (أقل من 18 ، 18 ، 19 ، فوق 19) ؛ مدة الحصول على DEUG من اجل ثلاث مستويات تصنيف (سنتان ، 3 سنوات ، 4 سنوات).في هذا المثال لدينا n = 797 : p = 3 ؛ c_1 = 2 ، c_2 = 4 ، c_3 = 3 ؛ c = 9. جدول بيرت المقابل يكون على الشكل التالي :

	Bac C	Bac D	<18	18 ans	19 ans	>19	2 ans	3 ans	4 ans
Bac C	583	0	108	323	114	38	324	192	67
Bac D	0	214	25	97	68	24	76	82	56
<18	108	25	133	0	0	0	84	35	14
18 ans	323	97	0	420	0	0	224	137	59
19 ans	114	68	0	0	182	0	73	75	34
>19	38	24	0	0	0	62	19	27	16
2 ans	324	76	84	224	73	19	400	0	0
3 ans	192	82	35	137	75	27	0	274	0
4 ans	67	56	14	59	34	16	0	0	123

IV-2-2 مبادئ ACM :

- **المشكلة :**

الهدف هو دراسة الروابط التي قد توجد بين المتغيرات p التي تم النظر فيها. في الواقع ، في حالة قياس اين يمكن تقديم البيانات في شكل جدول بيرت ، تركيب لجداول الاقتران ، يتم النظر فقط في الروابط بين

المتغيرات التي يتم أخذها مثنى مثنى (والتي تسمى في الإحصائيات تفاعلات من الرتبة الثانية). لدراسة هذه الروابط، فان النهج المتبع سيكون من نفس الطبيعة كما في AFC.

- **الطريقة**

تتمثل ACM ببساطة من تطبيق AFC لجدول Burt المعتبر . نستطيع لتوضيح أن هذا الأمر منطقي من ناحية ، ومن ناحية أخرى ، في حالة وجود AFC الخاص بـجدول بيرت المتعلق بمتغيرين نوعيين (الحالة حيث p = 2) ، نتحصل تقريبا على نفس النتائج حالة جداول الاقتران المتعلق بهذين المتغيرين: ACM اذن هو تعميم ل AFC.

مثال توضيحيي

يتعلق هذا المثال بالطلاب المسجلين لأول مرة في جامعة العلوم الاجتماعية لمدينة تولوز (Toulouse I) في خريف عام 1990 ، في السنة الأولى من DEUG في القانون ، ويتبع حتى عام 1996.

- **البيانات**

هناك 1635 طالبًا تم أخذهم في الاعتبار (n = 1635) و 5 متغيرات نوعية (p= 5). المتغيرات هي :
- الجنس (SEX)، له مستويين : ذكر ، انثى ؛
- سلسلة البكالوريا (Série de bac) ، بـ 5 مستويات تصنيف :bacA ، bacB ، bacCouD، bacG، autbac؛
- سن الحصول على البكالوريا (l'âge d'obtention du bac) ، بثلاث مستويات: 18. ، 19. ، 20.
- الفئة الاجتماعية والمهنية للآباء (la catégorie socio-professionnelle)، مع 6 مستويات تصنيف: art + com (الحرفيون والتجار), empl (الموظفون)، inter (المهن الوسيطة)، ouvr (العمال)، Prolib (المهن الحرة), autcsp غيرها من CSP ؛
- النجاح (La réussite) ، على الأقل في DEUG ، مع مستووين من التصنيف : نعم ، لا.

يتم تقديم البيانات في شكل ملف مكون من 1635 صفًا و 5 أعمدة، وهو معطى تحت الأسطر الثلاثة الأولى و الثلاثة الأخيرة.

1	4	3	2	2
1	4	3	2	2
2	1	3	1	1
...	...	...	...	...
1	3	3	2	2
2	5	3	5	2
1	2	2	2	2

ملاحظة 4 يجب هنا ملاحظة الخصوصية التي هي ، عمليًا ، منهجية تقريبًا مع هذا النوع من البيانات (العديد من المتغيرات النوعية): مستويات التصنيف لكل متغير قد تم ترميزه ب 1،2 ... هذا، بالطبع، أكثر ملاءمة لتسجيل البيانات على دعم الكمبيوتر. لكن هذا يتطلب إعادة ترميز لجعل ملف مستويات التصنيف الأولية في جدول أو على الرسم البياني. في الواقع ، إذا تمكنا من فهم ،في الملف أعلاه ، أن

الرقم "2" الذي يظهر في الصف 3 والعمود 1 يمثل فتاة ، بينما يمثل ملف"2" الذي يظهر في السطر 1 والعمود 4 يمثل ابنًا لموظف ، ولن يكون هذا ممكنًا عندما سيواجه "2" في الرسم البياني. وبالتالي، تكون مرحلة إعادة تشفير البيانات عامة ضرورية قبل تنفيذ ACM.

IV-2-3 المعطيات في طريقة ACM

كما هو الحال بالنسبة للفصول الاخرى فان عملية معالجة هذه البيانات قد تمت باستخدام برنامج Excel-stat.

ـ لوحة بيرت

النتيجة الأولى المقدمة هي جدول بيرت ، ويسمى دائمًا "جدول الاقتران" في Excelstat. بالطبع، القراءة أكثر تعقيدًا من قراءة جدول الاقتران العادي الذي يزاوج بين متغيرين فقط. عندما نفسر ارتباطًا بين متغيرين (من بين جميع المتغيرات التي تم أخذها في الاعتبار) ، فانه يُنصح عمومًا بمعاينة جدول بيرت لقراءة التكرارات المقابلة (من الضروري ان نتأكد دائمًا من أننا لا نفكر في الرقم الضعيف جدًا). لاحظ أن الأرقام الهامشية (هي نفسها في الصفوف والأعمدة لأن جدول بيرت متماثل)

ليس من السهل تفسيرها هنا: كل منها يساوي الى عدد مستوى التصنيف المقابل مضروب في عدد p من المتغيرات التي تم النظر فيها (هنا 5). أخيرًا، إجمالي القيم يساوي العدد من الملاحظاتn مضروبة في p^2 هنا (p=25) ، أي 40875.

	Fille	gars	autbac	bacA	bacB	bacCouD	bacG	.18.	.19.	.20.	art+com	autcsp	empl	Inter	ouvr	prolib	non	oui
fille	1014	0	32	366	339	92	185	508	321	185	106	232	99	156	143	278	550	264
gars	0	621	19	126	285	94	124	221	210	190	61	119	54	98	74	215	390	231
autbac	32	19	51	0	0	0	0	6	9	36	2	20	4	6	10	9	45	6
bacA	366	126	0	492	0	0	0	255	167	70	56	107	47	70	57	155	287	205
bacB	339	258	0	0	597	0	0	314	190	93	62	91	69	120	78	177	265	332
bacCouD	92	94	0	0	0	186	0	117	54	15	15	24	6	21	9	111	70	116
bacG	185	124	0	0	0	0	309	37	111	161	32	109	27	37	63	41	273	36
.18.	508	221	6	255	314	117	37	729	0	0	63	125	61	132	90	258	311	418
.19.	321	210	9	167	190	54	111	0	531	0	65	115	63	74	62	152	326	205
.20.	185	190	36	70	93	15	161	0	0	375	39	111	29	48	65	83	303	72
art+com	106	61	2	56	62	15	32	63	65	39	167	0	0	0	0	0	97	70
autcsp	232	119	20	107	91	24	109	125	115	111	0	361	0	0	0	0	233	118
empl	99	54	4	47	69	6	27	61	63	29	0	0	153	0	0	0	87	66
inter	156	98	6	70	120	21	37	132	74	48	0	0	0	254	0	0	143	111
ouvr	143	74	10	57	78	9	63	90	62	65	0	0	0	0	217	0	143	74
prolib	278	215	9	155	177	111	41	258	152	83	0	0	0	0	0	493	237	356
non	550	390	45	287	256	70	273	311	326	303	97	233	87	143	143	237	940	0

oui	464	231	6	205	332	116	36	418	205	72	70	118	66	111	74	256	0	695

- نسب الجمود الذاتي للأبعاد المختلفة

يعطي الجدول التالي القيم الذاتية ، أو القصور الذاتي وفقًا للمحاور (القصور الذاتي الرئيسي او ما يعرف بالعطالة الكلية) ، تفكيك مربع كاي على المحاور ونسب القصور الذاتي التي يتم إرجاعها بواسطة كل محور

القيم الذاتية ونسبة العطالة المفسرة

	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13
Valeur propre	1,000	0,990	0,983	0,944	0,932	0,917	0,866	0,847	0,792	0,792	0,770	0,757	0,733
Inertie (%)	7,115	7,045	6,997	6,719	6,630	6,522	6,160	6,024	5,632	5,632	5,478	5,383	5,217

المشكلة هي أن هذا الجدول لا يمكن تفسيره مثل الجداول المماثلة الموجودة في ACP و AFC، حيث يحتوي جدول Burt على الكثير من المعلومات الزائدة عن الحاجة (على وجه الخصوص، متناظرة وجميع الأرقام تتكرر مرتين). النسب المئوية أعلاه المتعلقة بمجموع المعلومات الواردة في الجدول، جاءت سيئة التقدير إلى حد كبيرمع نسبة خطأ كبيرة. وبالتالي ، فإن المحورين الأولين لهذا التحليل لا يمثلان 24.19النسبة المئوية للتشتت الكلي (14.45 + 9.74) ولكن أكثر. للأسف لا يمكننا أن نعرف ما هي النسبة الفعلية. لذلك يجب أن تؤخذ هذه النسب المئوية فقط كمؤشر.

4-2-IV احداثيات مستويات التصنيف ومساهماتها في القصور الذاتي

يتم أخذ مجموعتين فقط من النتائج في الاعتبار هنا: إحداثيات الأعمدة على المحاور ، مما يجعل من الممكن إنتاج الرسم البياني (الرسوم البيانية) ، اعتمادًا على عدد المحاور المحددة (اثنان أو أكثر) ؛ مساهمات الأعمدة في القصور الذاتي (التشتت) على طول كل محور والتي يمكن تفسيرها تمامًا كما هو الحال في AFC. الكميات الأخرى المستخدمة في AFC (المساهمات في khi اثنان ، الملفات الشخصية وجيب التمام المربّع) لم يعد لهما تفسير مباشر في ACM ولم يعدا في عموما لا تستخدم

ملاحظة : لأن جدول بيرت متماثل ، فإن صفوفه وأعمدته متطابقة .ال ذلك فإن عناصر ACM المتعلقة بالصفوف متطابقة مع العناصر المتعلقة بالأعمدة وبواسطة لذلك ، لم يتم توفيرها .

نعطي أدناه إحداثيات مجموعة مستويات التصنيف على المحورين الأولين (من أجل التبسيط ، سنستخدم هنا المحورين الأولين فقط) ، ثم مساهماتهم في العطالة من اجل كل من هذه المحاور.

يتم إعطاء الرسم البياني لمجموعة الإجراءات وفقًا للبعدين الأولين بواسطة شكل 1

Coordonnées des colonnes

	Dim 1	Dim 2
fille	-0.11125	-0.53743
gars	-0.11125	0.87754
autbac	1.62701	0.56575
bacA	-0.21630	-0.81059
bacB	-0.40520	0.09334
bacCouD	-0.40520	1.55368
bacG	-0.40520	0.08171
.18.	-0.68841	-0.11547
.19.	0.09059	-0.16661
.20.	1.21001	0.46039
art+com	0.05265	-0.36354
autcsp	0.65135	-0.25675
empl	-0.02064	-0.68415
inter	-0.22781	-0.14436
ouvr	0.51077	-0.29683
prolib	-0.58262	-0.29683
non	-0.58262	-0.00691
oui	-0.77603	0.00935

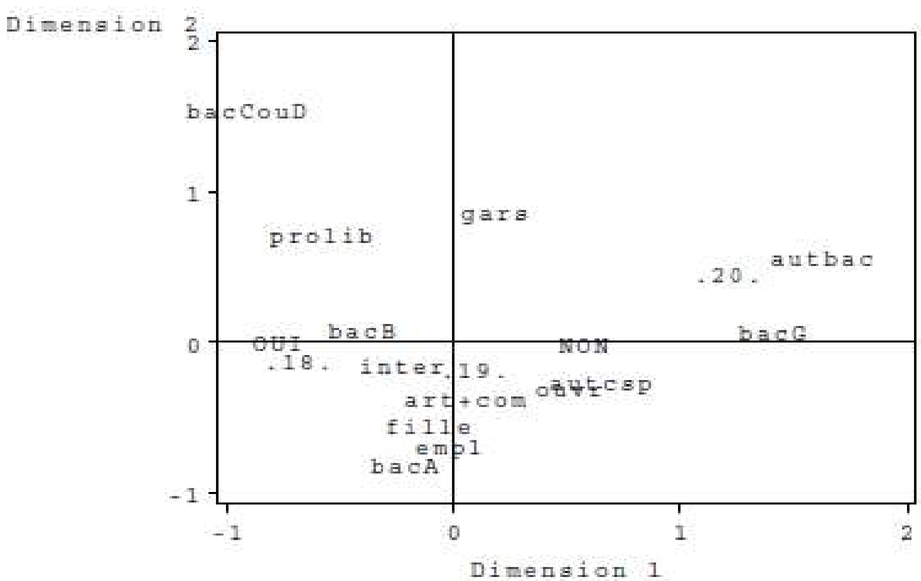
Contribution partial dans l’inertie pour les points des colonnes

	Dim 1	Dim 2
fille	0.004087	0.141475
gars	0.006674	0.231007
autbac	0.043970	0.007885
bacA	0.007497	0.156160
bacB	0.031923	0.002513
bacCouD	0.050491	0.216889
bacG	0.199587	0.000997
.18.	0.112521	0.004695
.19.	0.001419	0.007120
.20.	0.178820	0.038396
art+com	0.000151	0.010661
autcsp	0.048500	0.011177
empl	0.000021	0.034593
inter	0.004293	0.002557
ouvr	0.018438	0.009236
prolib	0.054504	0.124588
non	0.100786	0.000022
oui	0.136315	0.000029

- **التفسير**

سنقوم فقط بتفسير البعدين الأولين: يكفي هنا ، وعلاوة على ذلك ،يتم تفسير أي بُعد آخر وفقًا لنفس المبدأ. المبدأ العام هو تحديد مستويات التصنيف التي لها مساهمات مهمة في المحاور ثم إلقاء نظرة على تحديد المواقع على الرسم البياني

التمثيل الرسومي حسب البعدين الأولين



في المحور الأول ، هذه المساهمات هي مساهمات البكالوريا G (تقريبًا 20٪) ، وحاملي البكالوريا اصحاب 20 سنة أو أكبر (حوالي 18٪) ، وكذلك أولئك الذين تبلغ أعمارهم 18 عامًا أو أقل (ما يزيد قليلاً عن 11٪) والنجاح أو الفشل (13.6٪ و 10٪ على الترتيب). من خلال مراقبة الرسم البياني، نرى ان هذا المحور 1 يميز بين النجاح على اليسار والفشل على اليمين. لذلك يمكننا تفسيره بشكل أساسي كمحور للمناظرة بين النجاح والفشل في DEUG في القانون. المستويات المحددة أعلاه (مساهمات قوية في المحور 1) وقريبة من الفشل في هي البكالوريا G , والحصول المتأخر على البكالوريا ؛ مستوى التصنيف القريب من النجاح هي الحصول على بكالوريا الشباب. لذلك يمكننا أن نرى أن العامل المهيمن في اجتياز هذا DEUG هو العمر الذي تم فيه الحصول على شهادة البكالوريا (بمعنى آخر ، جودة التعليم الثانوي). بالإضافة إلى ذلك ، يبدو أن البكالوريا G غير مناسبة لدراسات القانون. في المحور 2 ، أهم المساهمات هي مساهمات الأولاد (أكثر بقليل من23٪) والفتيات (أكثر من 14٪ بقليل) ، البكالوريا C أو D(21.7)٪) ، البكالوريا A (15.6٪) ،المهن الحرة (حوالي 12.5٪). ما زلنا نلاحظ تمييزا واضحا على طول المحور 2 بين الأولاد أعلاه والفتيات أدناه. غالبًا ما يحمل الأولاد شهادة البكالوريا G أو D وغالبا ما يكون الوالدان ينتميان للمهن الحرة بينما البنات هن غالبًا ما يكون حاملي البكالوريا A، دون أن يرتبط ذلك بوضوح بالنجاح أو الفشل. وهي ما تمثل ظاهرة ملحوظة بشكل واضح في التعليم الثانوي والتي نجدها هنا نتيجة لتحليلنا.

ملاحظة 6: أخيرًا، نلاحظ الخصوصية التالية: في ACM، كل المتغيرات التي تؤخذ في الاعتبار تلعب، بديهيا، نفس الدور: لا يمكن للتحليل امتياز أي منها. ومع ذلك، من الناحية العملية، من الشائع أن يلعب المتغير دورًا محددًا في ذلك هذا ما نحاول شرحه من الآخرين: هذا بالضبط هو الحال بالنسبة للمتغير "النجاح" في المثال أعلاه. يظهر هذا الدور المحدد، ربما، فقط على مستوى التفسير بعبارة أخرى لاحقا. في هذه الحالة فهذا يعني ، بطريقة اخرى ، أن ACM عملت بشكل جيد ، أي أن المتغيرات تشرح الظاهرة (هنا متغير "النجاح") تم أخذها في الاعتبار وتم تسليط الضوء عليها من خلال التحليل.

3-IV المساعدة على تفسير النتائج

كما هو الحال في "ACP"، يتم تقدير نوعين من الحسابات المساعدة على التفسير والتي تسمح بفهم للرسوم البيانية التي تمنحها طريقة «AFC»:

- مساهمات الاصناف أو مستويات التصنيف في التباين أو العطالة,
- جودة تمثيل مستويات التصنيف.

- **مساهمات مستويات التصنيف في التباين**

التباين لاحد المكونات الرئيسية والذي يعرف ب C_i^r ، يساوي القيمة الذاتية ذات الرتبة r . يمكن حساب هذا التباين كما تم التطرق اليه سابقا في الدرس.

من أجل كل من المتغيرين النوعيين، فإن مجموع مساهمات مستويات التصنيف يساوي الى 100 ٪ . كما أن مستويات التصنيف التي تساهم بقوة في تباين أحد المكونات الرئيسية فإنها تشرح أو تعطي تفسير لهذا المكون الرئيسي.

- **جودة تمثيل الطرائق:**

يتم تمثيل كل مستوى i من مستويات التصنيف بواسطة نقطة في فضاء $m2$. كما في " ACP" ، يمكننا قياس جودة تمثيل هذه النقطة على المحور r . كما يتم الاستعانة ب بكل الحسابات المتعلقة ب مركز السحابة , مربع المسافة ل i الى ذلك المركز ومربع التجب والذي يعبر عن دودة التمثيل للنقطة I.

يتم تفسير الطرائق (مستويات التصنيف) الممثلة بشكل جيد من خلال المكون الرئيسي .بالنسبة لفرد معين، فإن مجموع جودة التمثيل لجميع المحاور يساوي 100٪.

4-IV دراسة حالة مع التطبيق :

تتعلق البيانات التالية بالمزارع (الحيازات) الفلاحية في منطقة من المناطق الساحلية. البيانات مأخوذة من "الجداول الاقتصادية في ميدي بيرينيه" التي نشرتها المديرية الإقليمية لتولوز INSEE ، في عام 1996 (البيانات المتعلقة بسنة 1993 ؛ أرقام مقربة إلى أقرب العشرات). تفصيل الحيازات الزراعية في منطقة ميدي بيرينيه وفقاً للقسم و S.A. U (في عام 1993). حيث تم تقسيم المزارع البالغ عددها 73000 مزرعة في منطقة ميدي بيرينيه في هذا الجدول الاقتران وفقا للقسم (في الخطوط ، 8 مستويات تصنيف) وS.A.U (المنطقة الزراعية المستخدمة في الأعمدة ، 6 مستويات تصنيف).

رموز القسم: ARIE = Ariège ; AVER= Aveyron ; H.G= haute- Garonne ; GERS= Gers ; LOT= Lot ; H.P= Hautes-Pyrénées ; TARN= Tarn ; T.G= Tarn-et- Garonne

رموز الفئات او الاصناف : أقل من 5 هكتار = INF05, ما بين 5 و 10 هكتارات = S0510 , ما بين 10 و 20 هكتار = S1020, ما بين 20 و 35 هكتار = S2035, ما بين 35 و 50 هكتار = S3550, أكثر من 50 هكتار = PLU50

نلاحظ أن المتغير الثاني ليس نوعيًا ولكنه كمي مستمر في الواقع، فإننا تعتبره هذه الطريقة نوعي، مما يعني أن الترتيب الطبيعي للصفوف لم يؤخذ على الإطلاق بعين الاعتبار. يمكننا دائمًا محاولة العثور على هذا الترتيب عند تفسير الرسم البياني.

données

	INF05	SO510	S1010	S2035	S3550	SUP50
ARIE	870	330	730	680	470	890
AVER	820	1260	2460	3330	2170	2960
H.G.	2290	1070	1420	1830	1260	2330
GERS	1650	890	1350	3540	2090	3230
LOT	1940	1130	1750	1660	770	1140
H.P.	2110	1170	1640	1500	550	430
TARN	1770	820	1260	2010	1680	2090
T.G.	1740	920	1560	2210	990	1240

فيما يلي سوف نقوم بتفسير مختلف النتائج الاساسية لطريقة ال AFC والتي تم الحصول عليها عن طريق المعالجة الاحصائية ببرنامج EXCEL STAT. ليكن في العلم ان هذه النتائج هي قابلة للمقارنة مع النتائج المقدمة عن طريق معظم البرامج الاحصائية الاخرى على غرار SPSS, R...S-PLUS89..

- النتيجة الاولى المقدمة من طرف البرنامج هي الجدول الاولي والجي يعرف بجدول الاقتران بهوامشه :

Tableau de contingence

	INF05	SO510	S1010	S2035	S3550	SUP50	SUM
ARIE	870,000	330,000	730,000	680,000	470,000	890,000	**3970**
AVER	820,000	1260,000	2460,000	3330,000	2170,000	2960,000	**13000**
H.G.	2290,000	1070,000	1420,000	1830,000	1260,000	2330,000	**10200**
GERS	1650,000	890,000	1350,000	3540,000	2090,000	3230,000	**11750**
LOT	1940,000	1130,000	1750,000	1660,000	770,000	1140,000	**8390**
H.P.	2110,000	1170,000	1640,000	1500,000	550,000	430,000	**7400**
TARN	1770,000	820,000	1260,000	2010,000	1680,000	2090,000	**9630**
T.G.	1740,000	920,000	1560,000	2210,000	990,000	1240,000	**8660**
SUM	**13190**	**7590**	**12170**	**15760**	**9980**	**14310**	**73000**

المساهمة في قيمة مربع كأي: النتيجة الثانية هي قيمة مؤشر خي مربع (5375.49) الذي نحصل عليه عن طريق جمع على جميع الخلايا - الخانات - لجدول الاقتران، للكميات:

$$\frac{(n_{ij} - \frac{n_{i.}n_{.j}}{n})^2}{\frac{n_{i.}n_{.j}}{n}}$$

في الواقع، يوفر جدول مساهمة مربع كأي الكميات المذكورة أعلاه في كل خلية .، مما يجعل من السهل الكشف عن الخلايا (وبعبارة أخرى تقاطعات القسم والسطح) الذي يساهم أكثر في مربع خي ، وبالتالي يساهم في تعريف العلاقة .

المساهمات في إجمالي إحصائيات مربع كأي

La contribution dans le totale de la valeur de khi deux

	INF05	**SO510**	**S1010**	**S2035**	**S3550**	**SUP50**	**SUM**
ARIE	32.5	16.6	7.02	36.59	9.75	16.05	**118.51**
AVER	995.17	6.21	39.54	97.62	86.79	66.49	**1291.82**
H.G.	108.42	0.08	46.26	62.87	12.97	54.64	**285.24**
GERS	106.4	90.05	189.25	0	145.61	372.82	**903.14**
LOT	118.62	76.11	88.22	12.64	123.92	154.86	**574.38**
H.P.	446.82	208.58	133.83	5.96	210.68	718.07	**1723.94**
TARN	0.52	32.81	74.33	2.29	100.34	21.67	**231.96**
T.G.	19.63	0.43	9.36	61.97	31.77	123.35	**246.51**
SUM	**1827.07**	**430.88**	**587.82**	**279.96**	**721.82**	**1527.95**	**5375.49**

لنعتبر على سبيل المثال الخلية (1.1) اي تقاطع كل من ARIEو INF05 : حيث يمكن ان نحصل على القيمة التقاطعية (32,5) عن طري تطبيق العلاقة السابقة :

$$\frac{(870 - \left(\frac{3970.13190}{73000}\right))^2}{\frac{3970.13190}{73000}} \approx 32,50$$

هذه القيمة منخفضة نسبيًا (مقارنة بالقيم الأخرى في الجدول)، مما يعني ذلك ان المزارع الصغيرة جدا أو الاقل استغلالا (أقل من 5 هكتارات) ليس لها شيء خاص جدا في Ariège .

دعونا ننظر الآن في الخلية (2،1) ، أي AVER x INF05 ؛ نحصل:

$$\frac{(820-(\frac{13000.13190}{73000}))^2}{\frac{13000.13190}{73000}} \approx 995.17$$

هذه القيمة هي الأكبر في جدول المساهمة ، مما يعني أنه في أفيرون ، بالذات تقدم المزارع الصغيرة خصوصية واضحة للغاية: فهي إما كثيرة جدًا ، إما عدد قليل جدًا (التربيع الذي يتدخل في التعبير عن مربع خي يزيل الاشارة لا يسمح بقول أي من الحالتين هو الذي يكون فعلا) .هذا هو جدول ملفات تعريف- الخطوط أدناه، الذي سيجعل من الممكن تبديد هذا الغموض: في حين أن هذا النوع من الاستغلال يمثلون ما بين 14٪ و 29٪ من جميع المزارع في الأقسام الأخرى وهي 6.3 ٪ فقط في أفيرون ، وبعبارة أخرى هي قليلة جدًا .هذه الظاهرة هي عنصر مكون مهم جدا للارتباط بين الأقسام والسطوح.

Test d'indépendance entre les lignes et les colonnes	
Khi² (Valeur observée)	**5590,525**
Khi² (Valeur critique)	49,802
DDL	35
p-value	< 0,0001
alpha	0,05

من خلال جدول اختبار الاستقلالية لكاي تربيع بين الاسطر والاعمدة أي بين متغيرات الدراسة يظهر لنا ان قيمة كاي تربيع المشاهدة اكبر بكثير من قيمة كأي تربيع النظرية 5590.525 >49.802 وهذا يقودنا الى قبول الفرضية البديلة اي هناك ارتباط بين مستويات تصنيف السطر ومستويات تصنيف العمود.

هذان الجدولان هما اللذان يعطيان ملفات تعريف الصف بالنسبة للأول وملفات تعريف الأعمدة بالنسبة للثاني. برنامج SPSS لا يعبر عنها بالنسب المئوية ، ولكن بالترددات او التكرارات، بحيث ان المجاميع (في الصفوف للأول وفي الأعمدة بالنسبة للثاني) تساوي الى الواحد (1).

Profils (lignes) :							
	INF05	**SO510**	**S1010**	**S2035**	**S3550**	**SUP50**	**Somme**
ARIE	0,219	0,083	0,184	0,171	0,118	0,224	1
AVER	0,063	0,097	0,189	0,256	0,167	0,228	1
H.G.	0,225	0,105	0,139	0,179	0,124	0,228	1
GERS	0,129	0,070	0,106	0,278	0,164	0,253	1
LOT	0,231	0,135	0,209	0,198	0,092	0,136	1
H.P.	0,285	0,158	0,222	0,203	0,074	0,058	1
TARN	0,184	0,085	0,131	0,209	0,174	0,217	1
T.G.	0,201	0,106	0,180	0,255	0,114	0,143	1
Moyenne	0,192	0,105	0,170	0,219	0,128	0,186	1

لقد أشرنا بالفعل أعلاه إلى اهمية الملفات في تحليل جدول الاقتران. من الواضح أن الاختلافات في ملفات التعريف ، من صف إلى آخر أو من عمود إلى آخر ، هي التي تقوم بتحديد العلاقة بين المتغيرين المدروسين .لذلك يجب تأخذ بعين الاعتبار كضرورة في تحليل هذا الرابط.

Profils (colonnes) :							
	INF05	**SO510**	**S1010**	**S2035**	**S3550**	**SUP50**	**Moyenne**
ARIE	0,066	0,043	0,060	0,041	0,047	0,062	0,053
AVER	0,062	0,166	0,202	0,199	0,217	0,207	0,176
H.G.	0,174	0,141	0,117	0,109	0,126	0,163	0,138
GERS	0,125	0,117	0,111	0,211	0,209	0,226	0,167
LOT	0,147	0,149	0,144	0,099	0,077	0,080	0,116
H.P.	0,160	0,154	0,135	0,089	0,055	0,030	0,104
TARN	0,134	0,108	0,104	0,120	0,168	0,146	0,130
T.G.	0,132	0,121	0,128	0,132	0,099	0,087	0,117
Somme	1	1	1	1	1	1	1

مفهوم الجمود في « AFC »

يتعلق الجدول التالي في مخرجات برنامج SPSS بمفهوم القصور او العطالة .قبل التفاصيل، سنحاول توضيح هذه الفكرة في سياق خاص لطريقة « AFC ». دعونا نتذكر أولاً أن فكرة القصور الذاتي، أو التشتت، أمر أساسي في الإحصائيات.

فهو يحملنا الى .مفهوم التباين في الحالة أحادية البعد (أي في حالة متغيرة واحدة), وقد لعب بالفعل دورًا مركزيًا في ACP. هذا لا يزال نفس الحال بالنسبة ل AFC غير ان عبارته لها معنى خاص (إنه يمثل مؤشر phi-deux ، وهذا يعني مربع كأي مقسومًا على n ، العدد الإجمالي للملاحظات).

كل هذا موضح أدناه ، ببساطة قدر الإمكان ... للأسف للقراء غير الرياضيين، لا يمكن لهذه التفسيرات الالتفاف حول تقنية رياضية معينة غير انه يمكن لهم الاحتفاظ فقط بالفكرة النهائية.

ملف تعريف الصف هو عنصر يحتوي على مصطلحات c (c هو عدد الأعمدة في الجدول التحليلي) الذي يكون مجموعه 1. من وجهة نظر رياضية ، يمكننا بالتالي تمثيل كل ملف تعريف خطي جانبي بواسطة شعاع في مساحة شعاعية ذات c بعد. إحداثيات هذا الشعاع هي تشكيلات ملف تعريف الصف المقابل.

وهكذا نحصل على مساحة من الاشعة r في المساحة المعنية (r هو عدد الخطوط من الجدول الذي تم تحليله)، يمكن تحديد مركز barycenter الخاص به ، أي نقطة الوسط (كل من هذه الإحداثيات barycenter هو المتوسط المرجح للإحداثيات المقابلة لمجموع خطوط التعريف الاسطر ؛ الأوزان هي الأعداد الهامشية للصفوف. مكننا أن نفعل نفس المنطق على ملفات تعريف الأعمدة.

- نسب الجمود للأبعاد المختلفة

كما هو الحال في ACP، فإن الجدول الذي يعطي حصة الجمود التي تم إرجاعها حسب كل بُعد (كل محور) يسمح بمعرفة الجودة الإجمالية للنتائج (ولا سيما الرسوم البيانية) عندما يحتفظ بعدين أو ثلاثة أبعاد فقط. في مثال الحيازات الزراعية ، هذا الجدول موضح أدناه :

Valeurs propres et pourcentages d'inertie

Inertie totale	**0,076**				
Axes	**F1**	**F2**	**F3**	**F4**	**F5**

Valeur propre	0,056	0,013	0,004	0,001	0,001
Inertie (%)	74,470	17,572	5,523	1,644	0,792
% cumulé	74,470	92,042	97,565	99,208	100,000

اجمالي القصور للسحابتين (تلك الموجودة في ملفات تعريف الصف وملفات تعريف الأعمدة) هي متطابقة وتتحلل بنفس الطريقة وفقًا لمحاور العوامل المختلفة (أو الرئيسية، أو المحاور الرئيسية للقصور) التي تم الحصول عليها في التحليل.

وبالتالي ، يوجد جدول واحد فقط من النتائج في العمود "العطالة الرئيسية" (القصور الذاتي الرئيسي ، وهذا يعني وفقًا للمحاور الرئيسية) ، يعطي قيم القصور التي استعادتها كل محور (هذا هو القصور الذاتي للسحابة ، أو ملفات تعريف الصفوف أو ملفات تعريف الأعمدة ، المسقطة على هذا المحور). مجموع هذه العطالة يساوي phi-deux هنا = (0.07364).

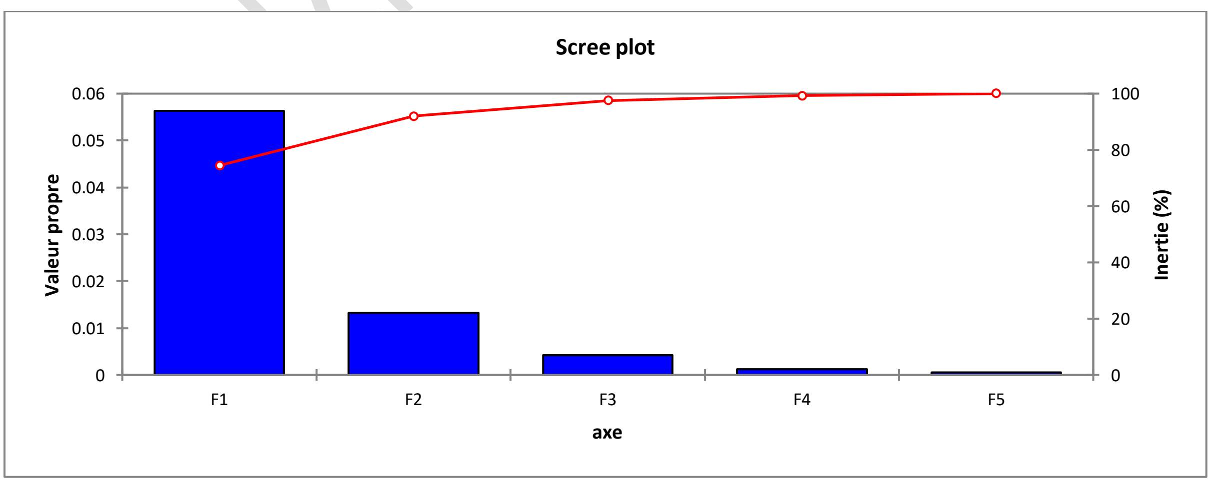
كما هو الحال في ACP، فان المحور الأول هو الذي يستعيد أكبر قدر من الجمود؛ ثم الثاني، بينما يكون متعامدًا مع الأول (في ما معنى مقياس مربع كاي)، بحيث يستعيد أيضا الحد الأقصى. و هكذا.

القيم المفردة ("القيم المفردة") ، الجذور التربيعية الإيجابية للجمود الرئيسي ، ليست ذات أهمية عملية ولا يتم استخدامها.

الكميات الموضحة في عمود مربعات "Chi-Squares " (مربع كأي) تساوي القصور الرئيسي. شفرات مضروبة في تكرار جدول الاقتران. لهذا السبب فان مجموعهم يساوي مربع خي دو (نذكر أن . $\chi^2 = n \times \phi^2$) يمكننا أيضا أن نعتبر أن كل محور من التحليل يستعيد جزءًا من مربع خي ، وبالتالي الارتباط بين المتغيرين الأوليين ، الأكثر مهم للمحور 1 وهكذا.

تمثل النسب المئوية ("Inertie (%)") نسب مربع كاي التي تم إرجاعها من قبل كل محور .كما هو الحال فيACP، نستخدم النسب المئوية التراكمية لاختيار البعد للاحتفاظ به. في مثالنا ، يمثل البعدين الأولين حوالي 95٪ من إجمالي العطالة الكلية. لذلك سنحتفظ فقط ببعدين ، مما سيسمح بإنتاج رسم بياني واحد فقط.

ملاحظة : عندما ننجز AFC لجدول الاقتران و الذي يشتمل على الصفوف و الاعمدة ، على سبيل المثال $r \geq c$ ، فان ابُعاد الفضاء الذي سوف توجد فيه مجموعة النتائج هو $c - 1$ (إذا كان لدينا $c \geq r$، فإن هذا البعد هو $r - 1$ ؛ بشكل عام ، فهو أقل من $\inf(r - 1, c - 1)$، وهكذا ، في المثال المذكور ، لدينا $r = 8$ و$c = 6$ ، وهو ما يفسر لماذا يوفر الجدول أعلاه فقط 5 أبعاد.



- **إحداثيات الصفوف والأعمدة**

هذه الإحداثيات هي التي تجعل من الممكن إنتاج الرسم البياني الذي يمثل في وقت واحد، وفقا للأبعاد 1 و 2 الاقسام و المساحات المستغلة . يتم تحديدهم وفقًا لذلك المبدأ الذي في ACP نعطي أدناه هذه الإحداثيات .يتم إعطاء الرسم البياني المقابل :

Coordonnées principales (lignes)					
	F1	**F2**	**F3**	**F4**	**F5**
ARIE	0,056	-0,110	-0,114	-0,039	-0,076
AVER	-0,241	0,185	-0,058	0,005	0,007
H.G.	0,041	-0,162	-0,060	-0,022	0,031
GERS	-0,265	-0,054	0,086	-0,026	0,007
LOT	0,262	0,042	-0,035	-0,020	0,008
H.P.	0,480	0,071	0,043	0,022	0,014
TARN	-0,093	-0,103	0,000	0,081	-0,010
T.G.	0,116	0,056	0,081	-0,016	-0,030

Coordonnées principales (colonnes)					
	F1	**F2**	**F3**	**F4**	**F5**
INF05	0,342	-0,170	0,025	0,003	-0,011
SO510	0,229	0,084	-0,027	0,010	0,065
S1010	0,159	0,157	-0,076	-0,006	-0,028
S2035	-0,101	0,076	0,100	-0,022	-0,002
S3550	-0,242	-0,011	0,001	0,081	-0,005
SUP50	-0,285	-0,103	-0,062	-0,034	0,006

يتم إعطاء تفسير الرسم البياني أدناه.

- **المساهمات في الجمود وفقا لكل محور**

لقد رأينا أن الجمود لكل سحابة (ملفات تعريف الصف وملفات تعريف الأعمدة) تم تقسيمها، بنفس الطريقة، وفقا للمحاور المختلفة. هنا، بما أننا نحتفظ فقط ببعدين فقط ، فإننا سوف نهتم فقط بالعطالة بالنسبة لهذين المحورين الأولين. لكل من المحورين المختارين ، تعطي الجداول أدناه حصص القصور المستحقة أولاً لكل صف (أو قسم) ، ثم لكل عمود (أو فئة SAU) هذه الحصة يتم التعبير عنها بالترددات وبالتالي المجموع يساوي 1.

Contributions (lignes)						
	Poids (relatif)	F1	F2	F3	F4	F5
ARIE	0,054	0,003	0,049	0,168	0,066	0,517
AVER	0,176	0,182	0,455	0,144	0,004	0,013
H.G.	0,138	0,004	0,273	0,119	0,053	0,216
GERS	0,172	0,215	0,038	0,307	0,093	0,014
LOT	0,113	0,139	0,015	0,033	0,036	0,011

H.P.	0,100	0,410	0,038	0,044	0,039	0,035
TARN	0,130	0,020	0,105	0,000	0,686	0,024
T.G.	0,117	0,028	0,028	0,185	0,023	0,171

Contributions (colonnes)						
	Poids (relatif)	**F1**	**F2**	**F3**	**F4**	**F5**
INF05	0,178	0,370	0,387	0,027	0,002	0,035
SO510	0,103	0,096	0,054	0,018	0,008	0,722
S1010	0,164	0,074	0,304	0,230	0,005	0,222
S2035	0,226	0,041	0,100	0,546	0,085	0,002
S3550	0,135	0,140	0,001	0,000	0,717	0,006
SUP50	0,193	0,279	0,153	0,179	0,183	0,012

كيف يتم تحديد هذه المساهمة؟ إذا أشرنا ب c_l^k لا حداثيات القسم رقم l (حيث $i = 1$ الى 8) على طول المحور k (k= 1,2) , العطالة على المحور k تساوي:

$$I_k = \sum_{i=1}^{r} \frac{n_{i.}}{n} (c_l^k)^2$$

وبالتالي فإن حصة القسم i تستحق:

$$\frac{\frac{n_{i.}}{n} (c_i^k)^2}{I_k}$$

.

لنأخذ مثال أفيرون (i = 2) على المحور 1 (k = 1). يوفر جدول القصور: $i_1 = 0.05501$, توفر الإحداثيات: $c_2^1 = -0.236684$, أخيرا ، جدول الاقتران الاولي يسمح لنا بكتابة $\frac{n_{2.}}{n} = \frac{13}{73}$
نستنتج أن مساهمة أفيرون في الجمود لسحابة الأقسام على طول المحور 1 تساوي

$$\frac{\frac{13}{73} \times (0,236684)^2}{0,05501} = 0,1813$$

القيمة الواردة في الجدول أعلاه.

يتم استخدام المساهمات في القصور الذاتي لتحديد الصفوف والأعمدة الأكثر أهمية في التحليل (وهذا يعني في تعريف الارتباط)، وعند الاقتضاء، تفسير محاور الرسوم البيانية.

دعونا نشير، مع ذلك، أنه في حالة AFC، فإن التفسير الملموس للمحاور ليس جوهريًا كما هو في ACP.يتم هذا التفسير فقط إذا كان من السهل القيام به وإذا كان ذلك سهلا في فهم النتائج .للقيام بذلك، نستخدم بالطبع الرسم البياني،

ولكن أيضًا مساهمات الصفوف وكذلك الأعمدة في القصور لسحابة النقاط. في المثال المعتبر يمكن بسهولة تفسير المحاور (ولا سيما الأول).

وهكذا يمكننا أن نرى أن أهم الأقسام في تعريف المحور 1 (تلك التي لها المساهمة الأكثر في الجمود) وهي Hautes-Pyrénéesو Gers Aveyron. . من وجهة نظر أحجامSAU ، وهذه المزارع تتمثل في المزارع الصغيرة جدا(INF05) ، الكبيرة جدا (SUP50) و والمعتبرة الحجم .(S3550).

يتم تحديد المحور 2، المتعلق بالأقسام، بشكل رئيسي من قبل Aveyron وHaute-Garonne ؛

بالنسبة لـ UAA ، فإن هذا يتعلق أساسًا بالمزارع الصغيرة جدًا وتلك التي تتراوح مساحتها بين 10و 20 هكتارا ، وبدرجة أقل ، مناطق S2035 و .SUP50 سنرى في النقطة 2.3 كيف تتدخل هذه العناصر في تفسير النتائج.

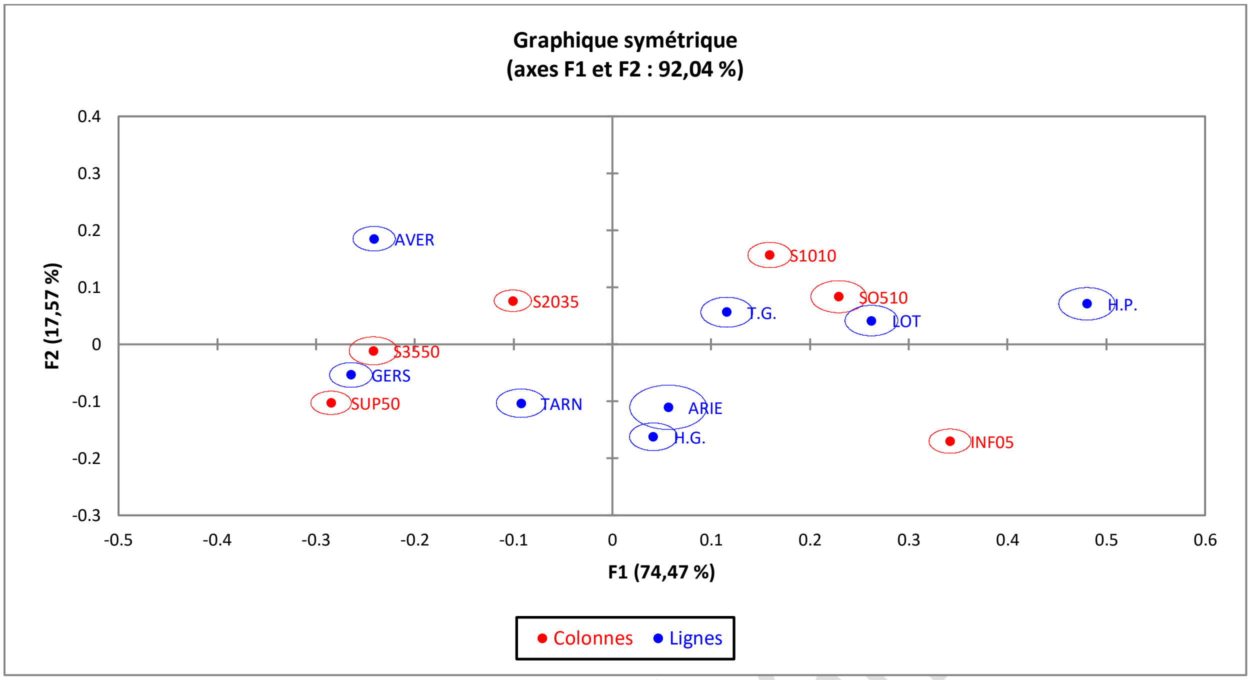
مربع جيب التمام

تشير هذه الكميات، كما هو الحال في ACP ، إلى جودة التمثيل في كل محور(بعبارة أخرى على كل بعد) لكل مستوى تصنيف (صف أو عمود). في كل من مساحات التمثيل لمستويات التصنيف (الصفوف و الأعمدة)، كل من أبعاد((inf (r – 1 ; c - 1 ، والزوايا التي نعتبرها جيب التمام هي الزوايا بين كل شعاع يمثل مستوى تصنيف والمحور المعتبر كلما كانت هذه الزاوية أصغر ،كلما اقترب جيب التمام (وبالتالي مربعه) من 1 ، كلما كانت جودة تمثيل الصورة أفضل لمستوى التصنيف على هذا المحور جيدة .كلما كبرت هذه الزاوية (قريبة من الزاوية المستقيمة) ، زاد جيب التمام (وبالتالي مربعه) قريب من 0 ، وكانت جودة تمثيل هذا المستوى من التصنيف على هذا المحور سيء.

نستخدم مربعات جيب التمام لأنه يمكننا إضافتها وفقًا للأبعاد المختلفة (خاصية هندسية كلاسيكية).

Cosinus carrés (lignes)					
	F1	**F2**	**F3**	**F4**	**F5**
ARIE	0,089	0,341	0,366	0,043	0,161
AVER	0,606	0,358	0,036	0,000	0,000
H.G.	0,052	0,796	0,109	0,014	0,028
GERS	0,864	0,036	0,092	0,008	0,001
LOT	0,953	0,024	0,017	0,005	0,001
H.P.	0,968	0,021	0,008	0,002	0,001
TARN	0,331	0,412	0,000	0,253	0,004
T.G.	0,552	0,131	0,272	0,010	0,036

Cosinus carrés (colonnes)					
	F1	**F2**	**F3**	**F4**	**F5**
INF05	0,798	0,197	0,004	0,000	0,001
SO510	0,813	0,108	0,011	0,001	0,065
S1010	0,448	0,434	0,103	0,001	0,014
S2035	0,383	0,220	0,379	0,018	0,000
S3550	0,897	0,002	0,000	0,101	0,000
SUP50	0,839	0,109	0,040	0,012	0,000



تفسير النتائج:

وتجدر الإشارة أولاً إلى أن هذا التفسير سيعتمد فقط على النتائج خلال البعدين فقط ، لأن 95٪ من المعلومات المفيدة (التي عبر عنها التشتت ، أي الجمود) الموجود في هذين البعدين فقط .سنرى أيضا أن الظواهر الأكثر مشاهدة هي المتعلقة او هي تلك التي كشف عنها البعد 1.

يقدم الشكل 2.2 المخطط الشريطي لمقاطع تعريف الخطوط (الأقسام) التي تسمح بذلك

فهم التفسيرات المتتالية بشكل أفضل (لاحظ أن المخطط الشريطي لملفات تعريف الأعمدة تحتوي على نفس المعلومات الإحصائية، لكن ملفات تعريف الصفوف تبدو أكثر ملاءمة، في هذا المثال، للمساعدة في التفسير).

بما أن المساحات المستغلة S.A.U مرتبة بشكل طبيعي ، فلنبدأ بدراسة تموضعها في المخطط (الخريطة) .أول شيء يمكن مشاهدته مباشرة هو أن ترتيبهم (تذكر، لم يؤخذ في الاعتبار في التحليل) يتم احترامه بشكل صارم على المحور 1 وهو بالتالي منظم للغاية: ترتب من اليمين إلى اليسار، السطوح، من الأصغر إلى الأكبر .لذلك ، بالإضافة إلى ذلك كلما تموضع القسم على اليمين ، وكلما كان لديه حيازات صغيرة والعكس صحيح.

وهكذا ، تتميز Hautes-Pyrénées بوجود العديد من المزارع الصغيرة والندرة النسبية للمزارع الكبيرة: 45٪ من المزارع أقل من 10هكتارات (المساحة التي تأتي مباشرة ، أقل من 37٪) ؛ فقط ما يزيد قليلا عن 13 ٪ لديهم أكثر من 35 هكتارًا (هنا مرة أخرى مساحة الأرض التي تأتي في الاخير، قريبة بالفعل من 23٪) .اذا الملف يعكس حقيقة أنه القسم الأكثر "جبلية" في المنطقة ، مثل اسمه

يشير إلى مكان آخر .من ناحية أخرى ، تتميز أفيرون وجيرز بوجود للمزارع الكبرى وندرة المزارع الصغيرة: تمثل المزارع التي تزيد مساحتها عن 35 هكتارًا تقريبًا 40٪ في أفيرون وأكثر من 45٪ في جيرز ؛ تمثل تلك التي تقل مساحتها عن 10 هكتارات فقط 16٪ في أفيرون و 21.6٪ في جيرز .الأسباب الجغرافية مختلفة: منطقة الهضاب ، والنعوش ، لأفيرون والسهول والتلال لجيرز ؛ في حالتين ، تفضل الجغرافيا وجود مزارع كبيرة.

يشار إلى أن جودة التمثيل في البعد 2 للأقسام المذكورة ممتازة. (اكثر من 0.99 لـ Aveyron و Gers و-Hautes Pyrénées؛ 0.97 للمساحة) ؛ نفس الشيء للأسطح المذكورة (0.99 لـINF05 ، 0.91لـS0510 ، 0.92لـ S3550 ، 0.97لـ SUP50).
وفيما يتعلق بمساهمات الأقسام في المحور 1 ، شهدت الأقسام الأربعة بانهم الوحيدون الذين لديهم مساهمات أكبر من 10٪ ، وهذا واضح جداً .نفس الشيء للأسطح INF05 و S3550 و SUP50 (S0510 أقل بقليل من 10٪). فيما يتعلق بالمساهمات في مربع كاي ، يمكننا التحقق من أن الظاهرة بالفعل تتوافق مع ما يقارب جميع المساهمات القوية (أكبر من 100).
السؤال التالي هو ما يمكن قوله أكثر .على وجه الخصوص، ماذا يمثل المحور 2؟ إن الحقيقة ليست واضحة للغاية ، وهي ظاهرة تابعة لما قيل في المحور 1 ، ليس من السهل تفسير الباقي .دعونا نحاول .بالنسبة للأقسام ، والمساهمات الوحيدة ذات الأهمية القليلة هي تلك هوت غارون و أفيرون ، التي تتعارض بوضوح على المحور 2. فيما يتعلق بالسطوح، المساهمات المهمة هي تلك INF05 وS1020 ، وبدرجة أقل ، S2035و .SUP50 وللغاية تم الإبلاغ عن عدد صغير من المزارع في أفيرون بمساحة أقل من 5 هكتارات (مساهمة قوية في مربع كاي). من ناحية أخرى، ينبغي أن نشير ، في هذا القسم ، إلى
عدد كبير من المزارع المتوسطة الحجم S.A.U بين 20 و 35 هكتار .هذا يسمح لذلك لتحسين صفة Aveyron الخاصة إلى حد ما: العديد من المزارع الكبيرة جدًا (SUP50) والمتوسطة (S2035) ؛ نسبة قريبة من المتوسط لإقليم السطوح S1020 و S3550 ؛ عدد قليل جدا من المزارع الصغيرة أقل من 10 هكتارات .ماذا عن هوت غارون؟ هو القسم الوحيد (مع Ariège ، ممثلة بشكل سيئ في الرسم البياني) تمتلك أكثر من 20٪ من المزارع التي تقل مساحتها عن 5 هكتارات وفي نفس الوقت أكثر من20 ٪ من المزارع التي تزيد مساحتها عن 50 هكتارا .وهي أيضًا قسم يوجد به عدد قليل نسبيًا من المزارع متوسطة الحجم .تأتي كل هذه الخصائص من موقعها الجغرافي ، الذي يمتد على طول المحور الشمالي والجنوبي ، مع الجنوب منطقة جبلية (Comminges) ، ومن الشمال ، منطقة السهول والتلال (سهل Garonneو Lauragais) .
في الختام، دعونا نحدد أننا قدمنا هنا، عن قصد، تفسيرا مفصلا للغاية لهذه AFCليس من الضروري دائمًا الخوض في الكثير من التفاصيل .سنحتفظ بالضروري - ومع ذلك ، يعتمد التفسير على الرسم البياني (الرسوم البيانية) ، يتطلب استخدام مختلف المؤشرات (المساهمات في المحاور ، مساهمات مربع كاي ، جيب التمام التربيعي) وهذا لا ينبغي أبدًا ان ننسى تحليل ملفات تعريف الصفوف والأعمدة) وبالتالي يجب أن ننظر إليها قبل تطوير أي عنصر من عناصر التفسير.

قائمة المراجع :

- **PHILIPPE C.,** 2006, « Principe de l'analyse factorielle », l'université de Versailles – St-Quentin Version, p 34.

- **CARPENTIER F-G.,** 2013/2014, « Analyse multidimensionnelle des données - Master 2ème année - Psychologie Sociale des Représentations », Réf : PSR92C – (polycopié et fichiers de données utilisés) / URL / http://geai.univ-brest.fr/~carpenti/

- **DUFOUR A.B., ROYER .M & LOBRY J.R.** 2017 « Fiche TD avec le logiciel : tdr 620 Initiation à l'analyse factorielle des correspondances », version 3.4.1, Page 2/16 URL / https://pbil.univ-lyon1.fr/R/pdf/tdr620.pdf

- **BACCINI A., SABATIER. P**., 2010, « Statistique Descriptive Multidimensionnelle (pour les nuls) », Institut de Mathématiques de Toulouse, UMR CNRS 5219, Toulouse cedex 9, France.

Université de , « Introduction à l'analyse factorielle des correspondances », 2000 **LAFFLY. D.,-** Pau, IRSAM, France, 13 page.

RAKOTOMALALA .R, 1989, « Analyse Factorielle des Correspondances », Université - Lumière Lyon 2, France, 58 pages.