محاضرات تحليل البيانات

1/ مراحل تحليل المعطيات

أ- الموضوع

ب- الإشكالية

ج- الفرضيات

د- استقراء الدراسات السابقة

ه- اختيار النموذج (الطريقة والمتغيرات)

و- جمع البيانات: – **طرق مباشرة** (الاستبيان المقابلة، الملاحظة)

– **طرق غير مباشرة** (مصادر ثانوية: كتب، دراسات، أوراق بحثية، واقع مختصة)

ز- التحليل الوصفي للبيانات

ح- التحليل النهائي (تقدير النموذج) وتفسير النتائج

ط- مقارنة النتائج المتوصل إليها مع الفرضيات (قبول أو رفض الفرضيات)

2/ أبرز أنواع النماذج المستخدمة في تحليل المعطيات:

أ- **نماذج سببية**: نماذج الانحدار، نماذج المعادلات الهيكلية، التكامل المتزامن، السببية

ب- نماذج متعددة الأبعاد: (ACP)، التحليل العاملي، تحليل المراسلات

3/ مفهوم تحليل المركبات الأساسية:

من النماذج الوصفية المتعددة الأبعاد تحليل مبدئي يعتمد على نموذج هندسي...

4 / مجال استخدام تحليل المعطيات

يتم الاستعانة بالتحليل العاملي من أجل عدة أغراض، لعل من أهمها:

- **التحليل العاملي الاستكشافي**: هو ذلك التحليل الذي يسمح باكتشاف وتلخيص نمط الارتباط بين المتغيرات محل الدراسة.

- تجميع المتغيرات الأولية في عوامل تشترك في الارتباط العالي فيما بينها.

- البعض الآخر يستخدم هذه الطريقة من أجل ابتكار طريقة لقياس بعد أو متغير، بتقدير البعد المراد قياسه كدالة (أو توليفة) خطية تربط المتغيرات الأولية فيما بينها.

- **التحليل العاملي التوكيدي**: ويستخدم حين تكون لدينا مثلا نظرية حول العناصر المختلفة في اختبار أو محور معين، فنبحث هل هذه العناصر تقع في عوامل متميزة تمثل تلك الكليات المحددة.

- مختصو القياس النفسي مثلا يستخدمون التحليل العامي في **بناء الاستبيان أو الاختبار**، حيث يستخدم التحليل العاملي لتطوير اختبار يقيس عدة أبعادا مختلفة.وفي هذه الحالة يبنى عادة اختبار من عدة عوامل كل عامل يحتوي مؤشرات فرعية لقياسه، ويستخدم التحليل العاملي لمعرفة هل المؤشرات لها ارتباط بالعامل بحيث تحذف وتغير المؤشرات التي لا ترتبط بالعامل، ويعاد جمع النتائج والتحليل حتى يتم الوصول إلى نتائج مرضية وبالتالي استخلاص اختبار ذو مصداقية.

كذلك يبحث بعض الباحثين في الاختلافات في العوامل بين المجموعات (مثلا الاختلافات بين اختبار أو استبيان أعد في ألمانيا وآخر في الجزائر، لتعديل الاختبار أو المقياس أو الاستبيان نلجأ إلى التحليل العاملي).

طريقة تحليل المركبات الأساسية (ACP)

(Analyse en Composantes Principales)

تحليل المركبات الأساسية (ACP: Analyse en Composantes Principales) ينتمي إلى مجموعة النماذج الوصفية المتعددة الأبعاد. [هذه النماذج التي ظهرت بداية سنوات الثلاثينات (fr) وتطورت سنوات الستينات]. وعلى اعتبار أنها طرق وصفية، فإن هذه الطرق من التحليل لا ترتكز على نماذج احتمالية بل تقوم على التحليل الهندسي.

طريقة تحليل المركبات الأساسية تقترح [انطلاقا من جدول للبيانات يضم (p) متغيرا كميا لوحدات أو أفراد عددها (n)]تمثيلات هندسية للمتغيرات والوحدات تبحث في العلاقات القائمة بينها. بحيث يسمح التمثيل الهندسي للوحدات بمعرفة مدى وجود شكل (أو هيكل) غير معلوم مسبقا لهذه المجموعة من الوحدات.

بشكل عام، تمثيل المتغيرات هندسيا يسمح بدراسة أشكال وهياكل العلاقات الخطية لكل المتغيرات محل الدراسة. وكذلك تبحث في إمكانية التمييز بين مجموعات من الوحدات التي تتماثل فيما بينها والتي تختلف وتتباين عن الوحدات الأخرى. [أما فيما يتعلق بالمتغيرات فنفرّق بين مجموعات المتغيرات المرتبطة فيما بينها وغير المرتبطة فيما بغيرها.]

وككل الطرق الوصفية، فإن القيام بالتحليل بطريقة ACP ليس أداة تحليل نهائية ولكنه يساعد على: 1\ فهم أفضل للبيانات التي نعمل عليها، 2\ كما يكون ACP وسيلة مساعدة على صياغة فرضيات الأبحاث والدراسات العلمية، 3\ تحديد المتغيرات الكامنة، 4\ تحسين جودة النموذج، 5\ أو توضيح لبعض نتائج التحليل وداعم لها،

جدول البيانات:

البيانات أو المعطيات هي قياسات في وحدات أو أفراد عددها n $\{u_1, u_2,\dots u_i,\dots u_n\}$ هذه القياسات هي شكل كتابة المتغيرات والتي يرمز لها بالرموز $\{v_1, v_2,\dots v_j,\dots v_p\}$. جدول البيانات الخام الذي سيبنى عليه التحليل يسمى X ويعطى بالصيغة التالية:

$$\mathbf{X} = \begin{array}{c} \\ u_1 \\ u_2 \\ . \\ u_i \\ . \\ u_n \end{array} \begin{array}{c} \begin{array}{cccccc} v_1 & v_2 & \dots & v_j & \dots & v_p \end{array} \\ \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1j} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2j} & \dots & x_{2p} \\ . & . & \dots & . & \dots & . \\ x_{i1} & x_{i2} & \dots & x_{ij} & \dots & x_{ip} \\ . & . & \dots & . & \dots & . \\ x_{n1} & x_{n2} & \dots & x_{nj} & \dots & x_{np} \end{bmatrix} \end{array} .$$

الوزن النسبي (الاحتمال): كل فرد يمكن أن يكون لـه وزن نسبي p. حيث ($p_1+p_2+\dots+p_n=1$)، وتبرز أهمية الوزن النسبي خاصة لما تختلف أوزان الأفراد النسبية (البيانات المجمعة)، وعادة ما نجد الوزن النسبي لكل فرد مساويا للنسبة ($p = 1/n$).

يمكننا كتابة كل فرد U_i في شكل شعاع يحتوي على قيم الفرد أو الوحدة لكل مقياس من مقاييس الفرد والمتمثلة في المتغيرات (وعددها p متغيرا)، هندسيا يمكن تمثيل كل فرد بشعاع $\overrightarrow{ou}$ من المستوي R^p أو بالنقطة U_i.

$${}^tU_i = [\ x_{i1} \quad x_{i2} \quad \dots \quad x_{ij} \quad \dots \quad x_{ip}\ ] \quad \text{ce qui donne} \quad U_i = \begin{bmatrix} x_{i1} \\ x_{i2} \\ . \\ x_{ij} \\ . \\ x_{ip} \end{bmatrix} .$$

بنفس الطريقة يمكن كتابة المتغيرات v_j في شكل شعاع مركّباته هي قيم المتغير لكل فرد من أفراد العينة أو المجتمع (وعددها n فردا)، هندسيا يمكن تمثيل كل متغير بشعاع $\overrightarrow{ov}$ أو بالنقطة V_j من المستوي R^n.

$$V_j = \begin{bmatrix} x_{1j} \\ x_{2j} \\ . \\ x_{ij} \\ . \\ x_{nj} \end{bmatrix} .$$

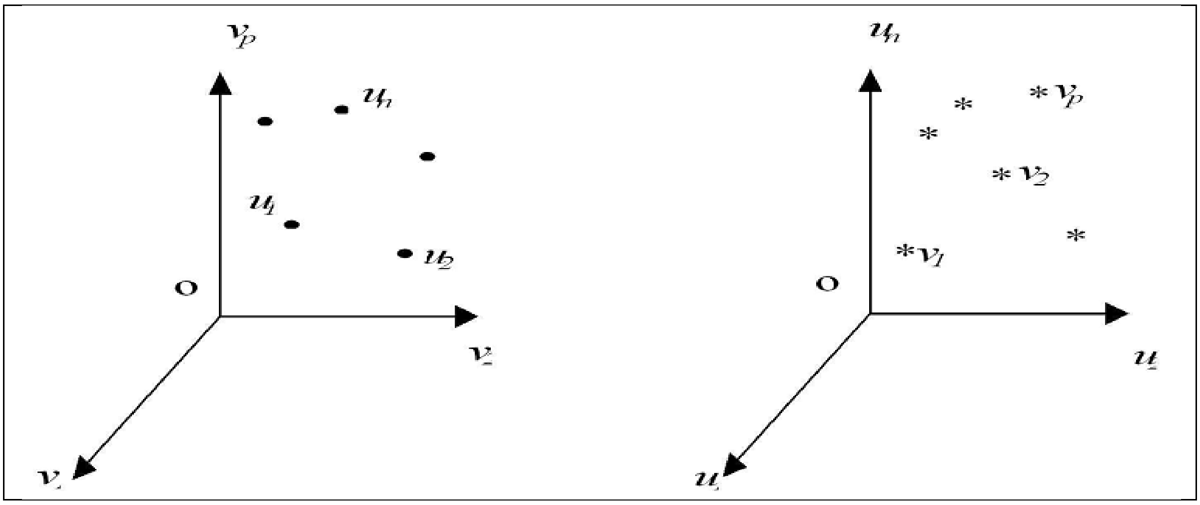
للحصول على صورة واضحة عن كل الوحدات يمكن تمثيلها على مستوي R^p ذو مركز معين $\boldsymbol{o}$ جميع بياناته معدومة. وتمثل كل فرد أو وحدة بنقطة من هذا المستوي، بحيث تعرف مجموعة النقط التي تمثل الأفراد بأنها "**سحابة الأفراد أو غيمة الأفراد**".

كما تعرف "**سحابة المتغيرات**" على أنها مجموعة النقط التي تمثل المتغيرات على المستوي R^n.

تقوم طريقة ACP على تمثيل المتغيرات و الأفراد بيانيا في معلم ثنائي أو ثلاثي الأبعاد أو أكثر من 3 أبعاد. لكن تجدر الإشارة إلى أنه في حالة وجود أكثر من 3 أبعاد يتعذّر الحصول على تمثيل هندسي للمتغيرات أو الأفراد.

الفكرة الأساسية لهذه الطريقة هي إيجاد نظام من المحاور يسمح بإعادة تموقع النقط بالنسبة إلى بعضها البعض بحيث تكون أقل تشوها وأكثر تناسقا قدر الإمكان، ليبدو جليا شكل العلاقة بين المتغيرات محل الدراسة.

سحابة المتغيرات   سحابة الأفراد



مثـال: تعطى بيانات الدخل Y، الإنفاق العام G، ومخزون رأس المال K، الواردات M، والصادرات X لدولة –أ– خلال الفترة (2008–2013).

جدول (1): متغيرات اقتصادية لدولة للفترة (2008، 2013)

مثال1						مثال 2					
Y	70	70	80	100	110	Y	70	70	80	100	110
G	1	2	4	5	10	M	50	55	70	90	95
KS	**4**	**6**	**8**	**10**	**11**	X	15	10	5	15	**10**

1– اكتب صيغة جدول المعطيات X.

2– بين صيغة شعاع بيانات الفرد الخامس (2013) للمثالين.

3– اكتب شعاع الاستهلاك، وشعاع الواردات.

1- تحديد المسافة:

لتمثيل نقاط هندسيا يجب حساب المسافة بين نقطتين من المعلم. المسافة الإقليدية الكلاسيكية في الفضاء R^p بين نقطتين u_i و $u_{i'}$ تعطى كالآتي:

$$\|\overrightarrow{U_i\, U_{i'}}\|^2 \qquad d^2(u_i, u_{i'}) = \sum_{j=1}^{p}(x_{ij} - x_{i'j})^2 .$$

في ظل هذا النمط المستخدم لقياس المسافة بين النقاط، كل المتغيرات تلعب نفس الدور، والمحاور المعرفة بالمتغيرات تكوّن معلما متعامدا. كما يمكن الحصول على الجداء الرقمي لشعاعين كما يلي (طويلة الشعاع حاصل الضرب):

$$\langle \overrightarrow{ou_i}, \overrightarrow{ou_{i'}} \rangle = \sum_{j=1}^{p} x_{ij}x_{i'j} = {}^tU_i\; U_{i'}$$

كما يتم حساب طوية الشعاع $(\overrightarrow{ou_i})$ بالقانون التالي:

$$\|\overrightarrow{ou_i}\|^2 = \sum_{j=1}^{p} x_{ij}^2 = {}^tU_i\; U_i$$

إذن يمكن تعريف الزاوية α بين الشعاعين من خلال قيمة الجيب تمام الخاص به:

$$\cos(\alpha) = \frac{\langle \overrightarrow{ou_i}, \overrightarrow{ou_{i'}} \rangle}{\|\overrightarrow{ou_i}\|\; \|\overrightarrow{ou_{i'}}\|} = \frac{\sum_{j=1}^{p} x_{ij}x_{i'j}}{\sqrt{\sum_{j=1}^{p} x_{ij}^2 \sum_{j=1}^{p} x_{i'j}^2}} = \frac{{}^tU_i\; U_{i'}}{\sqrt{({}^tU_i\; U_i)\; ({}^tU_{i'}\; U_{i'})}}.$$

مثال: نفس معطيات المثال السابق، يُطلب ما يلي في المثالين:

1- احسب طويلة الشعاع $(\overrightarrow{ou_3})$ ثم احسب المسافة بين (u_1) و (u_3) موضحا محاور المعلم الذي تمثل عليه النقط. $d(u_{1,}u_3)=11.18$، $\|\overrightarrow{ou_3}\|=80.5$، $\|\overrightarrow{ou_3}\|^2=6480$

مثال 2: $d(u_{1,}u_3\ )=95.78$، $\|ou_3\|=106.41$، $\|ou_3\|^2=11325$

2- ما هي الزاوية بين الشعاعين $(\overrightarrow{ou_1})$ و $(\overrightarrow{ou_3})$؟ $\cos\alpha_{13}=0.998,\ \alpha=3.62 \to 0$

مثال 2: $\cos\alpha_{13}=0.987\ ,\ \alpha=9.12 \to 0$.

2- اختيار المركز:

النقطة o التي ترتبط بشعاع المعطيات المعدومة ليست بالضرورة المركز المثالي، لأنه إذا كانت قيم معطيات أو بيانات نقاط سحابة الأفراد كبيرة، فإن السحابة ستبتعد عن المركز o. لذلك يبدو أكثر منطقية اختيار مركز مرتبط بالسحابة نفسها: وهو مركز ثقل الغيمة.

لتعريف مركز ثقل الغيمة G، يتم اختيار نظام ترجيح للأفراد (أو الوحدات).

لكل فرد وزن نسبي p_i. بحيث $(p_1+p_2+\dots+p_n=1)$، وخاصة لما تختلف أوزان الأفراد النسبية (البيانات المجمعة)، لذا في طريقة ACP عادة ما نجد الوزن النسبي لكل فرد مساويا للنسبة $(p=1/n)$.

$$\sum_{i=1}^{n} p_i\,\overrightarrow{Gu_i} = \overrightarrow{0}$$

مركز ثقل السحابة G هو إذن النقطة التي تكون فيها المعطيات هي القيم المتوقعة أو المتوسطة للمتغيرات:

$$G = \begin{pmatrix} \frac{1}{n}\sum_{i=1}^{n} x_{i1} \\ \vdots \\ \frac{1}{n}\sum_{i=1}^{n} x_{ij} \\ \vdots \\ \frac{1}{n}\sum_{i=1}^{n} x_{ip} \end{pmatrix} = \begin{pmatrix} x_{\bullet 1} \\ \vdots \\ x_{\bullet j} \\ \vdots \\ x_{\bullet p} \end{pmatrix} = \begin{pmatrix} \overline{V}_1 \\ \vdots \\ \overline{V}_j \\ \vdots \\ \overline{V}_p \end{pmatrix}$$

أخذ G كمركز، كما في الشكل (2)، يحولنا إلى العمل على جدول البيانات المركزة.

$$\mathbf{X}_c = \begin{bmatrix} x_{11}-x_{\bullet 1} & \cdots & x_{1j}-x_{\bullet j} & \cdots & x_{1p}-x_{\bullet p} \\ \vdots & & . & & \vdots \\ x_{i1}-x_{\bullet 1} & \cdots & x_{ij}-x_{\bullet j} & \cdots & x_{ip}-x_{\bullet p} \\ \vdots & & \vdots & & \vdots \\ x_{n1}-x_{\bullet 1} & \cdots & x_{nj}-x_{\bullet j} & \cdots & x_{np}-x_{\bullet p} \end{bmatrix}$$

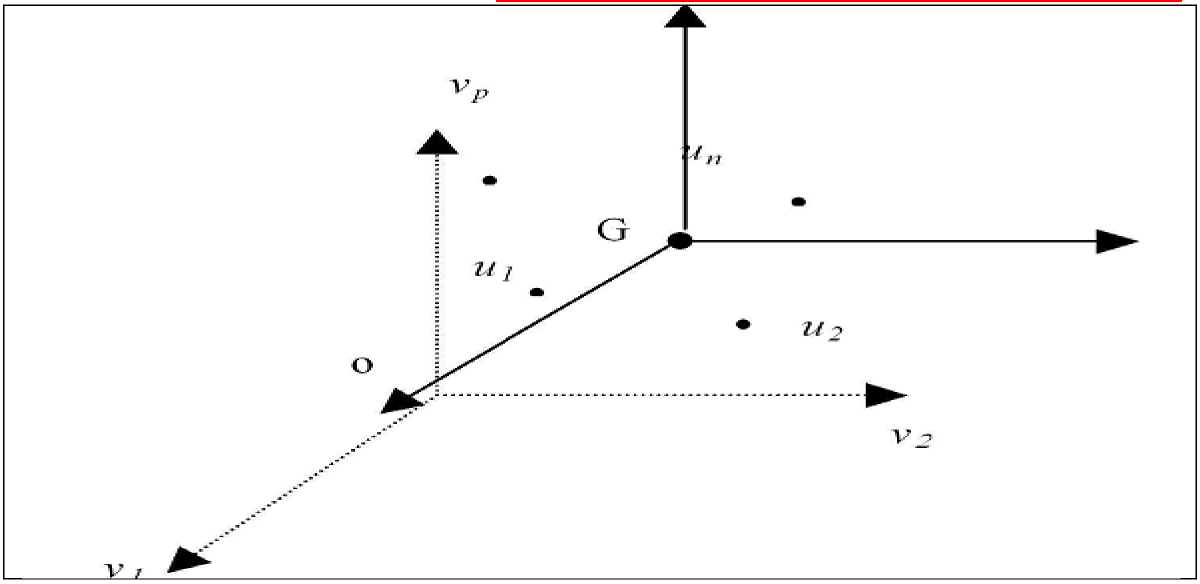
وشعاع المعطيات المركزة للفرد u_i هي:

$$U_{ci} = \begin{bmatrix} x_{i1}-x_{\bullet 1} \\ x_{i2}-x_{\bullet 2} \\ \vdots \\ x_{ij}-x_{\bullet j} \\ \vdots \\ x_{ip}-x_{\bullet p} \end{bmatrix}$$

في حين يعطى شعاع المعطيات المركزة للمتغير V_i هي:

$$V_{cj} = \begin{bmatrix} x_{1j} - x_{\bullet j} \\ \vdots \\ x_{ij} - x_{\bullet j} \\ \vdots \\ x_{nj} - x_{\bullet j} \end{bmatrix}.$$

الشكل (2): تمثيل سحابة الأفراد بعد تغيير المركز:



مثال: اكتب جدول البيانات المركزة X_C للمثالين السابقين.

3- عزم عطالة (Moments d'inertie) سحابة لأفراد (التباين الكلي):

أ- عزم العطالة الكلي لسحابة لأفراد بالنسبة لمركز الثقل G :

يعبر عن مدى تشتت سحابة الأفراد عن مركز ثقلها، فإن كان كبيرا كانت السحابة متشتتة أو متباعدة جدا وإن كانت صغيرة كانت السحابة مركزة جدا حول مركز ثقلها.

$$I_G = \frac{1}{n}\sum_{i=1}^{n} d^2(G, u_i) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{p}(x_{ij} - x_{\bullet j})^2 = \frac{1}{n}\sum_{i=1}^{n} {}^tU_{ci}\, U_{ci}\,.$$

وبقلب إشارات الجمع تصبح الصيغة كما يلي:

$$I_G = \sum_{j=1}^{p}\left[\frac{1}{n}\sum_{i=1}^{n}(x_{ij} - x_{\bullet j})^2\right] = \sum_{j=1}^{p}\mathrm{Var}(v_j)$$

حيث $\mathrm{Var}(V_j)$ تباين المتغير V_j. وفقا لهذه الصيغة يتضح أن قيمة عزم العطالة I_G تساوي أثر مصفوفة التباين والتباين المشترك $\sum$ للمتغيرات V_j التي عددها p.

$$I_G = \text{trace}\,(\Sigma)$$

$$\Sigma = \begin{pmatrix} C_{11} & C_{12} & \cdots & C_{1p} \\ C_{21} & C_{22} & \cdots & C_{2p} \\ \vdots & \vdots & & \vdots \\ C_{p1} & C_{p2} & \cdots & C_{pp} \end{pmatrix}$$

ملاحظة: تباينات المتغيرات غير متجانسة. وبالتالي فإن مساهمة المتغيرات في عزم العطالة يختلف باختلاف تباين المتغير، وأكثر متغير مساهمة في عزم العطالة هو أكبر المتغيرات تباينا.

مثال: نفس معطيات المثال السابق، يطلب ما يلي:

1- احسب مصفوفة التباين والتباين المشترك Σ.

$$\Sigma = \begin{pmatrix} 264 & 47.6 & 39.2 \\ 47.6 & 9.84 & 7.28 \\ 39.2 & 7.28 & 6.56 \end{pmatrix}$$

2- استنتج عزم العطالة الكلي، ثم بيّن أكثر المتغيرات تفسيرا لعزم العطالة الكلي.

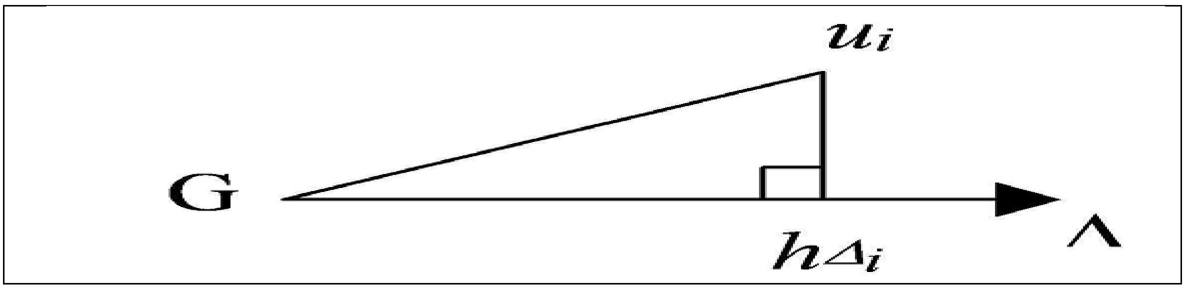
$I_G = tr\,\Sigma = 280.4$ ، المتغير الأول أي Y هو الأكثر تفسيرا ل I_G.

ب- عزم العطالة لسحابة الأفراد بالنسبة لمحور يعبر بمركز ثقل السحابة:

عزم العطالة لسحابة الأفراد بالنسبة لمحور Δ يعبر بمركز ثقل السحابة G يساوي بالتعريف إلى:

$$I_\Delta = \frac{1}{n}\sum_{i=1}^{n} d^2(h_{\Delta i}, u_i)$$

حيث $h_{\Delta i}$ هي الإسقاط العمودي ل u_i على المحور Δ. عزم العطالة في هذه الحالة يقيس قرب المحور Δ من سحابة الأفراد.



ج- عزم العطالة لسحابة الأفراد بالنسبة لفضاء جزئي V يعبر مركز ثقل السحابة G:

عزم العطالة لسحابة الأفراد بالنسبة لفضاء جزئي V يعبر مركز ثقل السحابة G يساوي بالتعريف إلى:

$$I_V = \frac{1}{n}\sum_{i=1}^{n} d^2(h_{Vi}, u_i)$$

حيث h_{Vi} هي الإسقاط العمودي ل u_i على الفضاء الجزئي V.

د- مكونات عزم العطالة لسحابة الأفراد:

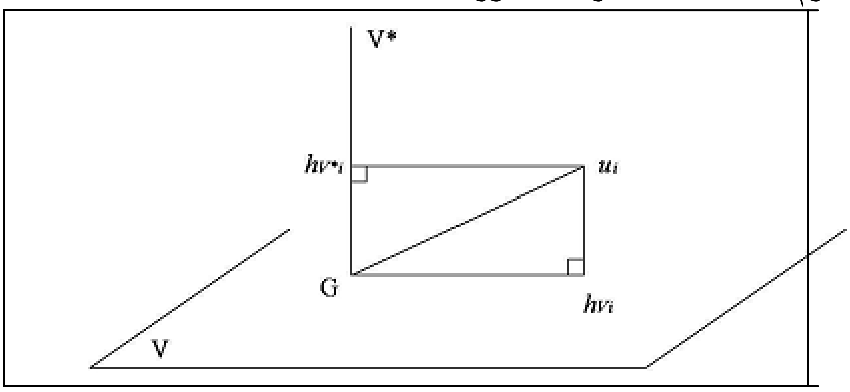
إذا كان الفضاء الجزئي V* المكملة العمودية للفضاء الجزئي V في المستوي R^p، و h_{V*i} الإسقاط العمودي للنقطة u_i على V*، بتطبيق نظرية فيثاغور (Pythagore)، نستطيع أن نكتب:

$$d^2(h_{Vi}, u_i) + d^2(h_{V^*i}, u_i) = d^2(G, u_i) = d^2(G, h_{Vi}) + d^2(G, h_{V^*i}).$$

إذن نستنتج العلاقة التالية المستنبطة من نظرية هيغنز (Huygens):

$$I_V + I_{V^*} = I_G$$

في الحالة الخاصة التي يكون فيها للفضاء الجزئي بعد (أو محور) واحد، I_{V^*} يمثل مقياس لاستطالة السحابة على طول هذا المحور. يسمى I_{V^*} "عزم العطالة المحمول بالمحور"، أو "عزم العطالة المفسر بالمحور".



عند إسقاط سحابة الأفراد على الفضاء الجزئي V يضيع عزم العطالة المقاس ب (I_V)، ولا يتم الاحتفاظ إلا بعزم العطالة المقاس ب (I_{V^*}).

إضافة إلى ذلك، إذا حللنا الفضاء R^p كمجموع للفضاءات الجزئية ذات البعد الواحد والمتعامدة فيما بينها:

$$\Delta_1 \oplus \Delta_2 \oplus \ldots \oplus \Delta_p$$

وبالتالي يمكن أن نكتب: $I_G = I_{\Delta_1^*} + I_{\Delta_2^*} + ... + I_{\Delta_p^*} \quad ...(*)$

أو بالصيغة: $I_G = \frac{1}{n}\sum_{i=1}^{n} d^2(G,u_i) = I_{\Delta_1^*} + I_{\Delta_2^*} + ... + I_{\Delta_p^*}$

4- البحث عن المحور Δ_1 الذي يمر عبر مركز ثقل السحابة G والذي يحقق عزم العطالة الأعلى

يتم البحث عن المحور Δ_1 الذي يمر عبر G مركز ثقل السحابة والذي يحقق عزم العطالة $I_{\Delta 1}$ الأعلى، وبالتالي إذا أردنا إسقاط السحابة على هذا المحور، فإنه سيشكّل الصورة الأكثر وضوحا والأقل تشوها للسحابة. باستخدام العلاقة بين عزوم العطالة كما في المعادلة (*) يقتضي البحث عن المحور Δ_1 الذي يحقق عزم العطالة $I_{\Delta 1}$ الأعلى البحث عن المحور Δ_1 الذي يعبر المركز G ويحقق عزم العطالة $I_{\Delta 1}$ الأقصى. [أي أنه المحور الذي تبتعد عنه النقط أكثر من بقية المحاور التي تعبر المركز G.]

يعرف المحور Δ_1 بالشعاع الإداري الوحدي $(\overrightarrow{G a_1})$. وبالتالي يجب إيجاد الشعاع $(\overrightarrow{Ga_1})$ بحيث يتحقق أقصى عزم عطالة $I_{\Delta 1}$ ممكن تحت قيد $(||\overrightarrow{Ga_1}||^2 = 1)$.

أ- الصيغة الجبرية لعزم العطالة $I_{\Delta 1}$ و $||\overrightarrow{Ga_1}||$:

$$d^2(G, h_{\Delta_1^* i}) = \left\langle \overrightarrow{Gu_i}, \overrightarrow{Ga_1} \right\rangle^2 = {}^t a_1 U_{ci}\, {}^t U_{ci}\, a_1$$

باستعمال الجداء الرقمي نستنتج أن:

$$I_{\Delta_1^*} = \frac{1}{n}\sum_{i=1}^{n} {}^t a_1 U_{ci}\, {}^t U_{ci}\, a_1 = {}^t a_1 \left[\frac{1}{n}\sum_{i=1}^{n} U_{ci}\, {}^t U_{ci}\right] a_1$$

المصفوفة الواقعة بين عارضتين هي مصفوفة التباين المشترك Σ للمتغيرات (V_j).

$$I_{\Delta_1^*} = {}^t a_1 \Sigma\, a_1$$

$$\left\|\overrightarrow{Ga_1}\right\|^2 = {}^t a_1\; a_1\ .$$

ب- البحث عن القيم القصوى:

المشكل الواجب حله هو إيجاد a_1 بحيث تأخذ الصيغة $({}^t a_1 \Sigma\, a_1)$ أقصى قيمة ممكنة تحت قيد ${}^t a_1\; a_1 = 1$، إذن إشكالية البحث عن قيمة مثلى لدالة متعددة

المتغيرات مرتبطة بقيد (المجاهيل المطلوب البحث عن قيمها هي قيم المصفوفة a_1). إذن يمكن استخدام طريقة مضاعف لاغرانج Lagrange.

في حالة البحث عن a_1، يجب حساب المشتقات الجزئية للصيغة $g(a_1)$:

$$g(a_1) = g(a_{11}, a_{12}, \ldots, a_{1p}) = {}^t a_1 \Sigma\ a_1 - \lambda_1({}^t a_1\ a_1 - 1).$$

باستخدام المشتقات المصفوفية نحصل على الصيغة التالية:

$$\frac{\partial g(a_1)}{\partial a_1} = 2\Sigma\ a_1 - 2\lambda_1\ a_1 = 0.$$

إذن يُطلب حل النظام التالي:

$$\begin{cases} \Sigma\ a_1 & - & \lambda_1\ a_1 & = & 0 & (1) \\ {}^t a_1\ a_1 & - & 1 & = & 0 & (2) \end{cases}$$

من المعادلة المصفوفية (1) في هذا النظام نستنتج أن a_1 هو شعاع المتجهات المميزة للمصفوفة Σ المرافقة للجذور المميزة λ_1. بضرب طرفي المعادلة المصفوفية (1) بـ${}^t a_1$ من جهة اليسار نجد أن: ${}^t a_1 \Sigma\ a_1 - \lambda_1\ {}^t a_1\ a_1 = 0$

بتعويض قيمة ${}^t a_1\ a_1$ بما يساويه من العادلة (2) في المعادلة الأخيرة نجد أن:

$${}^t a_1 \Sigma\ a_1 = \lambda_1\ .$$

إذا علمت أن الحد الأول للمعادلة السابقة يساوي عزم العطالة $I_{\Delta 1}$ الذي يجب أن يأخذ أقصى قيمة ممكنة. هذا يعني أن قيمة الجذر المميز λ_1 هي أكبر قيمة مميزة أو جذر مميز لمصفوفة التباين والتباين المشترك Σ، وبأن قيمة الجذر المميز هذا يساوي عزم العطالة المحمول بالمحور Δ_1.

عزم عطالة سحابة الأفراد للمحور Δ_1 يأخذ أعلى قيمة، وهذا المحور Δ_1 له شعاع إداري وحدي يحدده أول متجه مميز مرافق لأكبر جذر مميز لمصفوفة التباين المشترك.

5- البحث عن المحاور الأخرى

يتم البحث عن المحور الثاني Δ_2 عمودي على المحور Δ_1 ويمر عبر G مركز ثقل السحابة كما أنه يحقق عزم العطالة $I_{\Delta 2}$ الأدنى. يعرّف المحور Δ_2 بالشعاع الوحدي الإداري $\overrightarrow{Ga_2}$ أو a_2. عزم عطالة سحابة الأفراد بالنسبة إلى مكمّلتها العمودية $I_{\Delta 2}$ تعطى بالصيغة التالية التي يجب أن تكون أقصى ما يمكن:

$$I_{\Delta_2^*} = {}^t a_2 \Sigma\ a_2,$$

$$^t a_2 \; a_2 = 1 \quad \text{et} \quad ^t a_2 \; a_1 = 0.$$

تعظيم المعادلة الأولى يتم تحت شرطين، الشرط الأول مماثل له في معادلة المحور الأول، أما الشرط الثاني فيعني أن المحور الثاني عمودي على المحور الأول.

مثال توضيحي: للتوضيح تعطى النقاط التالية G (4.2) a_1 (4.3) و a_2 (3.2).

1- يطلب تمثيل هذه النقاط في مستوي ثنائي الأبعاد.

2- احسب إحداثياتها في المعلم الذي مركزه G.

3- أثبت أن المحورين الحاملين للشعاعين $\overrightarrow{Ga_1}$ و $\overrightarrow{Ga_2}$ متعامدان.

- بتطبيق طريقة مضاعفات لاغرانج تحت قيدين في هذه الحالة، نجد أن a_2 تمثل الشعاع (المتجه) المميز المرافق لثاني أكبر جذر مميز لمصفوفة التباين المشترك Σ.

ويمكن البرهان على أن المستوي المعرف بالمحورين Δ_2 و Δ_1 هو الفضاء الجزئي ثنائي الأبعاد الذي له أعظم قيمة عزم عطالة لسحابة الأفراد.

وبنفس الطريقة يمكن إيجاد المحاور الموالية والتي أشعتها الإدارية الوحدية تعرّف على التوالي بالأشعة (أو المتجهات) المميزة للمصفوفة Σ المرافقة للجذور المميزة ل Σ (من ثالث أكبر قيمة إلى أصغر قيمة للجذور المميزة λ).

إذن مصفوفة التباين المشترك Σ مصفوفة حقيقية متناظرة لها p شعاع (متّجها) مميزا حقيقيا تشكل مستوي متعامد المحاور R^p.

$$\left\{\begin{array}{ccccccc}
\Delta_1 & \perp & \Delta_2 & \perp & \dots & \perp & \Delta_p \\
a_1 & \perp & a_2 & \perp & \dots & \perp & a_p \\
\lambda_1 & \geq & \lambda_2 & \geq & \dots & \geq & \lambda_p \\
I_{\Delta_1^*} & \geq & I_{\Delta_2^*} & \geq & \dots & \geq & I_{\Delta_p^*}
\end{array}\right.$$

إذن بذلك ننتقل من المستوي المتعامد الابتدائي للمتغيرات المركزة إلى المستوي المتعامد الجديد للأشعة المميزة ل Σ. نسمي المحاور الجديدة بالمحاور (أو المركبات) الأساسية.

ملاحظة: عمليا آخر قيمة ل λ تكون غالبا معدومة بفعل تناقص عزم العطالة تدريجيا، لذلك يصبح لدينا أقل من p محور، مثلا (p-1) أو أقل.

6- مساهمة المحاور في عزم العطالة الكلي

استنادا إلى نظرية هيغنز يمكن تقسيم عزم العطالة الكلي لسحابة الأفراد إلى مكونات جزئية: $I_G = I_{\Delta_1^*} + I_{\Delta_2^*} + ... + I_{\Delta_p^*} = \lambda_1 + \lambda_2 + ... + \lambda_p$

المساهمة المطلقة للمحور Δ_k في I_G لسحابة الأفراد تساوي الجذر المميز الموافق له:

$$\mathrm{ca}(\Delta_k / I_G) = \lambda_k$$

المساهمة النسبية للمحور Δ_k في I_G لسحابة الأفراد تساوي:

$$\mathrm{cr}(\Delta_k / I_G) = \frac{\lambda_k}{\lambda_1 + \lambda_2 + ... + \lambda_p}$$

نستخدم عادة مصطلح ''نسبة عزم العطالة المفسر من طرف المحور Δ_k'' وينطبق هذا المفهوم على كل الفضاءات الجزئية التي تكوّنها المحاور الجديدة، وبالتالي يعرف عزم العطالة المفسر من طرف الفضاء الجزئي ذو البعدين (Δ_1، Δ_2) تساوي:

$$\mathrm{cr}(\Delta_1 \oplus \Delta_2 / I_G) = \frac{\lambda_1 + \lambda_2}{\lambda_1 + \lambda_2 + ... + \lambda_p}$$

وهذه النسب عبارة عن مؤشرات لمدى تغاير أو تشتت سحابة الأفراد المفسرة من طرف الفضاءات الجزئية.

7- تحديد عدد المحاور المستخدمة في التحليل:

أ\ معيار المساهمات النسبية للمحاور:

إذا كانت قيم آخر الجذور المميزة صغيرة نتجاهل التغاير الذي تفسره هذه المحاور. لذا عادة ما تمثل سحابة الأفراد على الفضاء الجزئي الذي تكوّنه d محورا الأولى في حالة ما إذا كانت نسبة مساهمة هذه المحاور في العزم الكلي تقترب من 1. مما يتيح اختزال التحليل إلى الفضاء الجزئي ذي d بعدا بحيث ($p > d$).

لذا يتم حساب مجموع المساهمات النسبية انطلاقا من أول محور، ونتوقف في تحديد عدد المحاور عند المحور الذي تكون فيه مجموع المساهمات النسبية: $\sum cr_j \leq 90\%$

في المثال المرفق: المحاور الثلاثة الأولى (Δ_1، Δ_2، Δ_3) تمثل مجتمعة 90.96% من I_G.

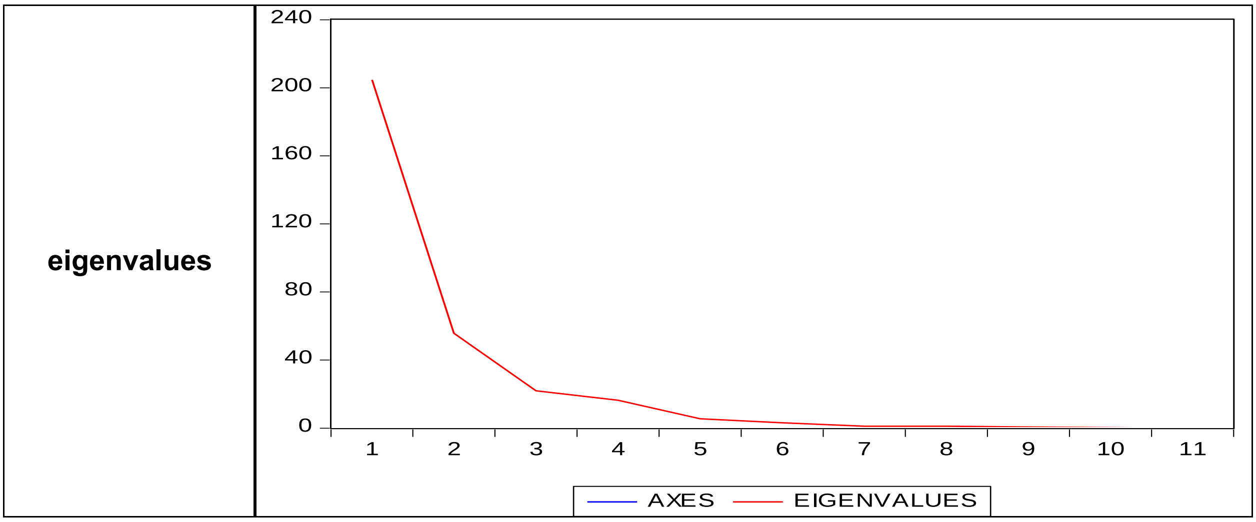
ب\ معيار الفرق الأول بين الجذور المميزة λ_j (المساهمات المطلقة):

يستخدم الفرق بين عزوم العطالة المتتالية لتحديد المحور الذي نتوقف عنده في تمثيل سحابة الأفراد والذي يختار عند ملاحظة انخفاض شديد في الفرق. في المثال المرفق نلاحظ انخفاضا شديدا في الفرق بين λ_4 و λ_5 بمقدار 10.88 وهو أكبر حتى من الفرق السابق، وبالتالي لا نأخذ بالمحورين 4 و5، ونتوقف عند المحاور الثلاثة الأولى.

	Eigenvalue	Difference	Proportion	Cumulative
AXE1	204.717	148.896	0.659161	0.65916
AXE2	55.822	33.853	0.179738	0.83890
AXE3	21.969	5.566	0.070736	0.90963
AXE4	16.402	10.880	0.052813	0.96245
AXE5	5.522	2.444	0.017780	0.98023
AXE6	3.078	1.966	0.009910	0.99014
AXE7	1.111	0.039	0.003578	0.99372
AXE8	1.072	0.440	0.003452	0.99717
AXE9	0.632	0.384	0.002034	0.99920
AXE10	0.248	0.248	0.000798	1.00000
AXE11	0.000	.	0.000000	1.00000

ج\ معيار اختبار الحصاة Scree Test (شكل):

نتوقف عند اختيار المركبة عند ملاحظة انخفاض كبير في الجذور المميزة.



آخر انخفاض كبير لقيم الجذور المميزة بين الجذر المميز 3 و 4، وبعدها تصبح الانخفاضات هامشية وبالتالي نتوقف في التحليل عند المحور 3.

المثال السابق (1): إذا علمت أن المساهمات النسبية للمحاور في I_G هي 65%، 25%، و10%. - يطلب تحديد عدد المحاور المناسبة للتحليل باستخدام المعايير الثلاثة.

8- تمثيل الأفراد في مستوى المحاور الجديدة

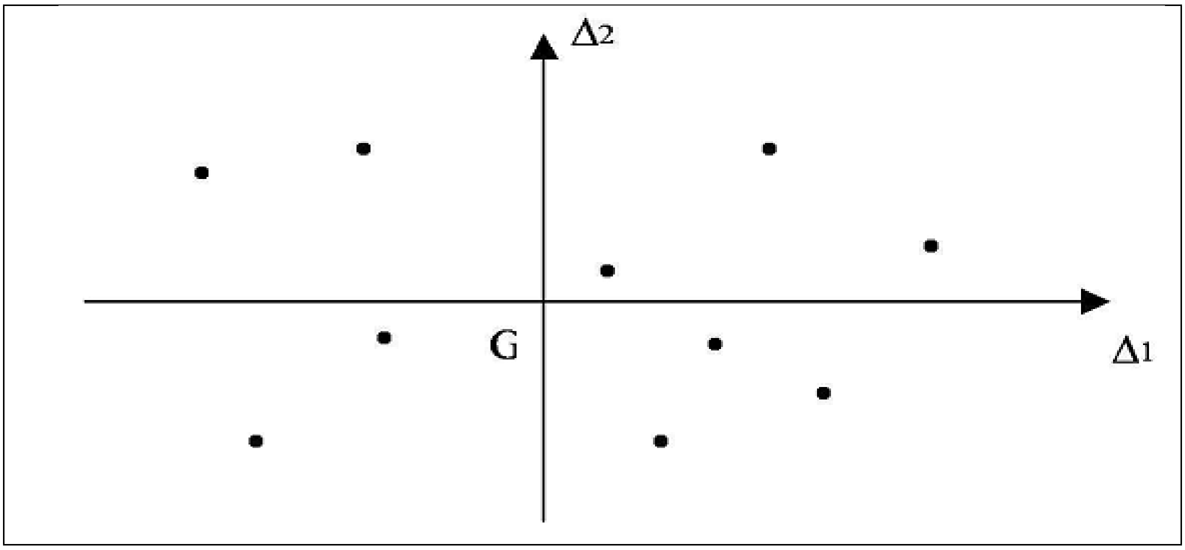
لتمثيل سحابة الأفراد على المستويات المعرفة بالمحاور الجديدة يكفي أن نحسب بيانات (أو معطيات) الأفراد بالنسبة إلى المحاور الجديدة. حيث للحصول على y_{ik} والتي تمثل بيانات الأفراد u على المحور Δ_k، نقوم بإسقاط عمودي للشعاع $\overrightarrow{Gu_i}$ على هذا المحور، فنجد أن:

$$y_{ik} = \left\langle \overrightarrow{Gu_i}, \overrightarrow{a_k} \right\rangle = {}^t a_k \, U_{ci}$$

و

$$Y_i = {}^t\mathbf{A} \, U_{ci}$$

بحيث y_{ik} شعاع بيانات الفرد u_i. وA مصفوفة التغيرات في المستوي (مصفوفة المتجهات المميزة المستقلة إحصائيا (أي أن منقولها يساوي مقلوبها) والتي طويلتها 1).



ملاحظة: اتجاه المحاور خاضع للتجربة والخطأ، لذا يعتبر اختياريا، بحيث يمكن أن نمثل المحاور في اتجاهات مختلفة مما يقلل من أهمية إشارة بيانات الأفراد. غير أنه يمكن تفسيرها بمعرفة موقع النقط من المحور (نقطتان متعاكستان بالنسبة إلى المحور).

أ- نوعية تمثيل الأفراد:

عندما تكون نقط الإسقاط متباعدة على محور (أو مستوي) فإن النقط التي تمثل هذه الوحدات (أو الأفراد) متباعدة في الفضاء، بينما تقارب الإسقاطات العمودية على محور أو فضاء لا يعني بالضرورة أن النقط قريبة فعليا في الفضاء.

أ- 1\ معيار جيب تمام الزاوية α_{ik}: لتفسير صحيح لقرب الإسقاطات العمودية لفردين على مستوي، يتعين التأكد من أن هذين الفردين ممثلان جيدا على المحور (أو المستوي أو الفضاء الجزئي) يجب أن تكون الزاوية بين الشعاع u_i G والمحور Δ_k (أو المستوي أو

$\longrightarrow$

الفضاء الجزئي) صغيرة. وللتأكد من ذلك يجب حساب جيب تمام هذه الزاوية. وكلما اقترب جيب تمام الزاوية من 1 كلما كانت الزاوية صغيرة وكلما اعتبرنا أن اتمثيل الفرد جيد.

$$\cos^2(\alpha_{ik}) = \frac{\left\langle \overrightarrow{Gu_i}, \overrightarrow{Ga_k} \right\rangle^2}{\left\| \overrightarrow{Gu_i} \right\|^2} = \frac{{}^t a_k U_{ci} \, {}^t U_{ci} \, a_k}{{}^t U_{ci} \, U_{ci}} = \frac{\left[\sum_{j=1}^{p} (x_{ij} - x_{\bullet j}) \, a_{kj} \right]^2}{\sum_{j=1}^{p} (x_{ij} - x_{\bullet j})^2}$$

استنادا إلى نظرية فيثاغور يمكن إثبات أن مربع جيب تمام الزاوية $\alpha_{ikk'}$ التي تقع بين الشعاع $\overrightarrow{Gu_i}$ والمستوي المحدد بالمحورين المتعامدين ($\Delta_k \oplus \Delta_{k'}$) يساوي إلى مجموع مربعي جيب تمام الزاويتين الواقعتين بين الشعاع وكل من المحورين على حدى:

$$\cos^2(\alpha_{ikk'}) = \cos^2(\alpha_{ik}) + \cos^2(\alpha_{ik'})$$

ويمكن تعميم القاعدة في حال وجود فضاء حزئي ذي d بعدا بحيث $(p > d)$.

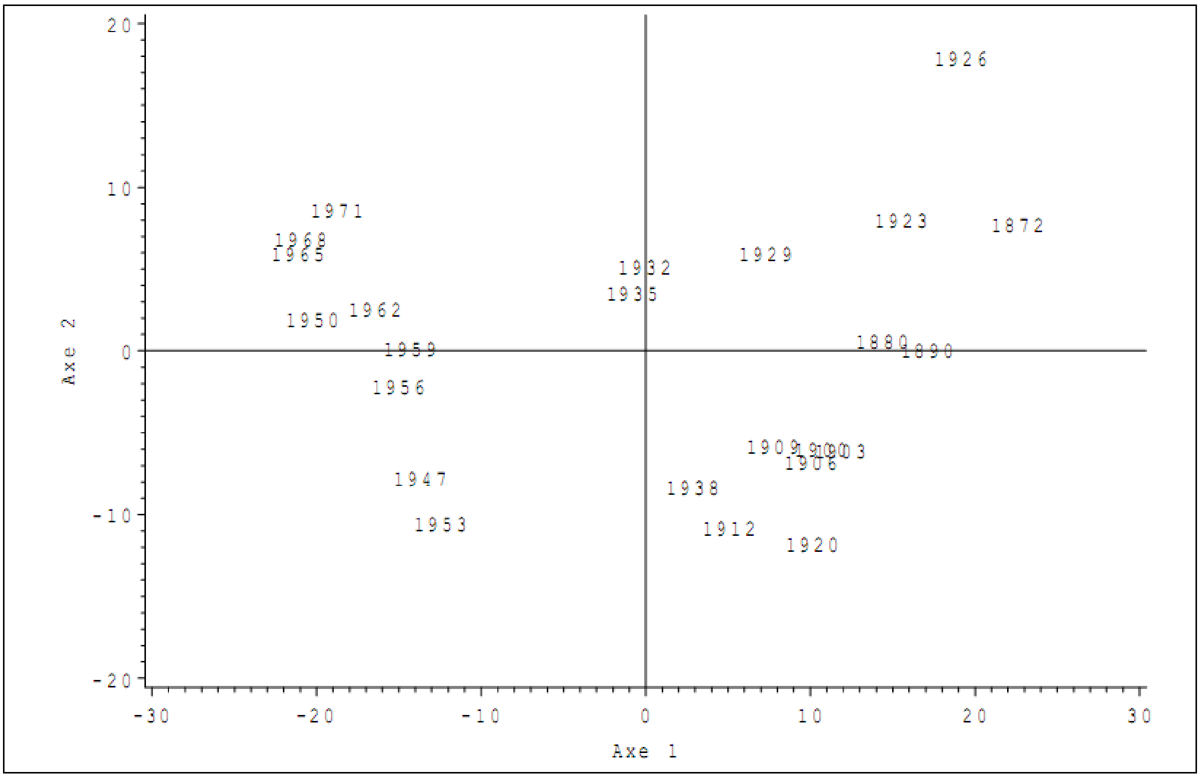
أ- 2\ معيار طويلة الشعاع $\|\overrightarrow{G\,ui}\|$: إذا كانت النقطة الممثلة للفرد قريبة من مركز ثقل سحابة الأفراد، كلما صغر مربع طويلة الشعاع $\overrightarrow{Gu_i}$، وكلما قلنا أن الفرد ممثل بطريقة جيدة على المحور Δ_k (أو المستوي أو الفضاء الجزئي).

<u>في المثال المرفق</u>: في جدول نوعية تمثيل الأفراد صفحة 31.

- احسب قيمة الزوايا في الفضاءين الجزئيين (1-2) والمعرّف بأول محورين أساسيين- و (1-2-3) لسنوات 1926 و 1932 و 1935 و 1971. ثم استنتج نوعية تمثيل هذه السنوات على هذين الفضاءين الجزئيين.

النتيجة: نجد أن أغلب السنوات ممثلة بطريقة جيدة ماعدا السنتين 1932 و 1935 غير الممثلتين بطريقة جيدة على الفضاء الجزئي (1-2). وبالتالي ظهورهما متقاربان على هذا المستوي كما في الشكل الموالي لا يمكن تفسيره مباشرة. وإذا عدنا إلى جدول بيانات الأفراد على المحاور الجديدة (أي المحاور الأساسية) نجد أن بيانات سنتي 1932 و 1935 مختلفة على المحور (Δ_3)، وبالتالي إسقاط الأفراد (السنوات) على الفضاء الجزئي (1- 2-3) يوضح التباعد الموجود بين النقطتين الممثلتين للسنتين 1932 و 1935.

<u>الإسقاطات العمودية لسحابة الأفراد على الفضاء الجزءي (1-2)</u>



ب- تفسير المحاور الجديدة كدالة إلى الأفراد:

عندما نحسب $I_{\Delta k}$ عزم العطالة المحمول بالمحور Δ_k، يمكننا إيجاد مقدار مساهمة الفرد u_i في هذا العزم (أو مقدار العزم الذي يعزى إلى الفرد u_i).

ب-1- المساهمة المطلقة لفرد في محور:

$$I_{\Delta^* k} = \frac{1}{n}\sum_{i=1}^{n} d^2(h_{\Delta_{ki}}, G).$$

تعطى المساهمة المطلقة للفرد u_i في عزم العطالة المحمول بالمحور Δ_k بالقانون التالي:

$$\mathrm{ca}(u_i/\Delta_k) = \frac{1}{n} d^2(h_{\Delta_{ki}}, G)$$

بما أن الوزن النسبي نفسه لكل الأفراد، ومن خلال القانون نجد أن فرد معين يساهم أكثر في عزم العطالة المحمول بمحور كلما ابتعد إسقاطه العمودي عن مركز ثقل سحابة الأفراد. وبالعكس تقل مساهمة الفرد كلما اقترب إسقاطه العمودي على المحور من مركز ثقل السحابة. تستخدم هذه المساهمات لتفسير المحاور الجديدة لتحليل المركبات الأساسية (ACP) كدالة للأفراد.

ب-2- المساهمة النسبية لفرد في محور:

بالنسة إلى لفرد معين u_i تعطى مساهمته النسبية في عزم العطالة المحمول على المحور بالصيغة التالية:

$$\mathrm{cr}(u_i/\Delta_k) = \frac{\frac{1}{n}d^2(h_{\Delta_{ki}}, G)}{I_{\Delta_k^*}} = \frac{\frac{1}{n}\left\langle \overrightarrow{Gu_i}, \overrightarrow{Ga_k} \right\rangle^2}{\lambda_k} = \frac{\frac{1}{n}\ {}^t a_k U_{ci}\ {}^t U_{ci}\ a_k}{\lambda_k}$$

اختبار هذه المساهمات النسبية يتيح تفسير المحاور الأساسية مع الأفراد. مع ملاحظة أن مجموع المساهمات النسبية للأفراد تساوي 1.

$$\sum_{i=1}^{n} \mathrm{cr}(u_i/\Delta_k) = 1.$$

<u>في المثال المرفق:</u> المساهمات النسبية للأفراد في المحاور موضحة في الجدول ص 31.

بحيث تمثل الخانة (Norme2)الطويلة ($\left\|\overrightarrow{Gu_i}\right\|^2$). ويمكننا ملاحظة أن أبعد السنوات عن مركز ثقل السحابة هي 1872، 1926، 1950، 1968.

نلاحظ كذلك أن السنوات في بداية الفترة (قبل سنة 1900) وفي نهاية الفترة (بعد 1960) بعيدة عن مركز ثقل السحابة مما يذكرنا بحدود الدراسة أو التحليل الذي يصلح فقط لهذه الفترة وبالتالي لا يسمح بأي استقراء أو تنبؤ مستقبلي.

يظهر من معطيات الأفراد على المحور الأول (ص 30) أنها تحترم بصورة شبه منتظمة الترتيب الزمني للأفراد، أي أن Δ_1 يمثل الزمن. أما المحور Δ_2 فأكثر فرد مساهمة فيه هو السنة 1926. على هذا المحور يبدو تناقض (تعاكس في الإشارة بين الأفراد) الفترة ما بين (1925-1935) و ما بعد 1960 من جهة مع الفترة قبل الحرب العالمية الأولى وفترة إعادة البناء التي أعقبت الحرب العالمية الثانية من جهة أخرى. إذن يمكننا الحديث عن تناقض بين فترات الأزمات الاقتصادية والفترات الأكثر نشاطا.

9- تمثيل المتغيرات

- يمكن التعامل مع تمثيل المتغيرات بطريقة مماثلة لتمثيل الأفراد. التمثيل سيتم على المستوي R^n وليس على المستوي R^p. وبغض النظر عن التماثل الشكلي بين تمثيل الأفراد والمتغيرات، يمكن استخدام الاختلاف المتعلق بكون المتغيرات ليس لها نفس معنى الأفراد.

أ\ تقدير مركبات المتغيرات الجديدة Z_j (تقدير المصفوفة Z)

نمثل الأفراد على فضاء تشكل محاوره المتغيرات القديمة، وقمنا بتغيير المستوي في هذا الفضاء، بحيث اعتبر أن هذه المحاور الجديدة هي توليفات خطية للمحاور القديمة. أي يمكن اعتبارها متغيرات جديدة تمثل توليفات خطية للمتغيرات القديمة. تسمى هذه المتغيرات الجديدة (Z_1، Z_2،....، Z_k،.... Z_p) بالمركبات الأساسية، بحيث Z_k المتغير الجديد المتعلق بالمحور Δ_k:

$$Z_K = \sum_{j=1}^{p} a_{kj} V_{cj} = X_C a_K \quad [X_{c(n.p)}\, a_{k(p.1)} = Z_k\, (n.1)]$$

بصفة عامة:

$$Z_{(n.P)} = [Z1\ Z2\ \dots.\ Zk\ \dots\ Zp] = X_{c\,(n.P)}\, A_{(P.P)}$$

ب\ مصفوفة التباين والتباين المشترك، ومعامل الارتباط للمتغيرات القديمة V_j والجديدة Z_j

$Var(X)= \sum(X^i-\overline{X})^2/n$ | $Cov(X,Y)= \sum(X_i-\overline{X})\,(Y_i-\overline{Y})/n$ | $R(X,Y)=Cov(X,Y)/\sigma_X\,\sigma_Y$

تعطى مصفوفة التباين والتباين المشترك $\sum$، ومعامل الارتباط للمركبات الأساسية فيما بينها أو مع المتغيرات الابتدائية وفقا للصيغ التالية:

تباين المتغيرات الجديدة Z_k

$$\mathrm{Var}(Z_k) = \frac{1}{n}\, \underset{1.p}{{}^t a_k}\, \underset{p.n}{{}^t\mathbf{X}_c}\, \underset{n.p}{\mathbf{X}_c}\, \underset{p.1}{a_k} = {}^t a_k\, \mathbf{\Sigma}\, a_k = \lambda_k$$

التباين المشترك بين V و Z

$$\mathrm{Cov}(Z_k, V_{cj}) = \frac{1}{n}\, \underset{1.p}{{}^t a_k}\, \underset{p.n}{{}^t\mathbf{X}_c}\, \underset{n.1}{V_{cj}} = \frac{1}{n}\, {}^t a_k\, {}^t\mathbf{X}_c\, \mathbf{X}_c \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix}$$

$$= {}^t a_k\, \mathbf{\Sigma} \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix} = \lambda_k\, {}^t a_k \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix} = \lambda_k\, a_{kj}$$

معامل الارتباط بين V و Z :

$$r(Z_k, V_{cj}) = \mathrm{Cor}(Z_k, V_{cj}) = \sqrt{\lambda_k}\frac{a_{kj}}{\sqrt{\mathrm{Var}(V_j)}}$$

بحيث a_{kj} هي الإحداثية رقم j للشعاع الإداري الوحدي a_k الخاص بالمحور Δ_k.

مصفوفة التباين والتباين المشترك للمركبات الأساسية Z (Σ_z): تعطى كالتالي:

$$\Sigma_Z = \frac{1}{n}\,{}^t\mathrm{A}\,{}^t\mathrm{X}_c\,\mathrm{X}_c\,\mathrm{A} = {}^t\mathrm{A}\,\Sigma\,\mathrm{A} = \Lambda$$

بحيث Λ هي المصفوفة القطرية للجذور المميزة للمصفوفة Σ.:

حسب نظرية الأطياف:

$$\Lambda = \begin{pmatrix} \lambda_1 & & (\mathbf{0}) \\ & \ddots & \\ (\mathbf{0}) & & \lambda_p \end{pmatrix}$$

استنتاج مصفوفة التباين والتباين المشترك للمركبات الأساسية Z والمتغيرات القديمة V:

$$\mathrm{Cov}(\mathbf{Z},\mathbf{V}) = \frac{1}{n}\,\underset{P.n}{{}^t\mathrm{X}_c}\;\underset{n.P}{\mathrm{X}_c}\;\underset{P.P}{\mathrm{A}} = \underset{P.P}{\Sigma}\;\underset{P.P}{\mathrm{A}} = \underset{P.P}{\mathrm{A}}\;\underset{P.P}{\Lambda}$$

إذا لاحظنا أن تباين متغير يساوي مربع طويلة الشعاع الذي يمثله وبأن معامل الارتباط بين متغيرين يساوي حاصل الضرب الرقمي للشعاعين الذين يمثلانهما، يمكننا تفسير زوايا الأشعة على أنها معاملات ارتباط.

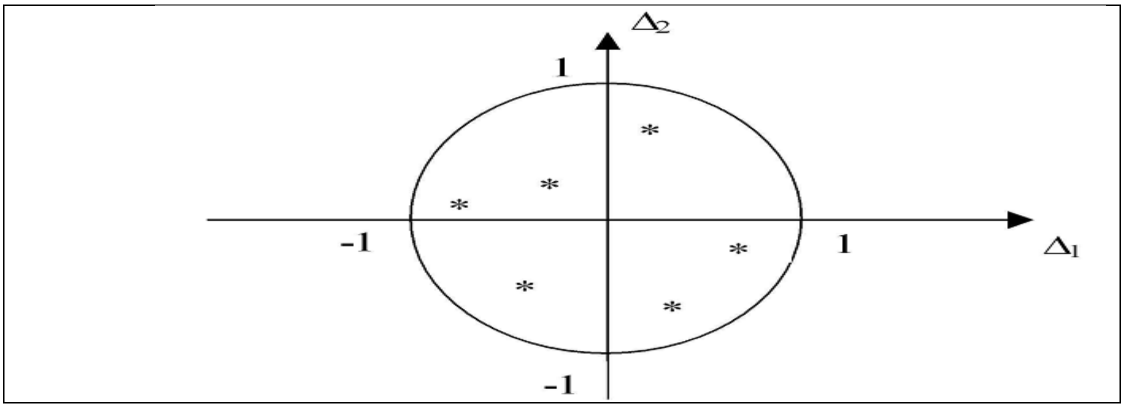
<u>في المثال المرفق</u>: معطيات (أو إحداثيات) المتغيرات الجديدة في مستوي الأفراد تقرأ من أعمدة جدول المعطيات (أو الإحداثيات) العاملية للأفراد (ص 30)، ومعاملات الارتباط بين المتغيرات الجديدة والمتغيرات القديمة (ص 32).

ج\ تحليل العلاقة بين المتغيرات القديمة V_j والجديدة Z_j

يتم تحليل العلاقة الخطية بين المتغيرات القديمة والجديدة إما بتحليل معاملات الارتباط والتباين المشترك بين هذه المتغيرات أو من خلال تحديد معامل الارتباط عند قيمة إسقاط النقط التي تمثل المتغيرات القديمة V_j (على دائرة الارتباط بين المتغيرات القديمة والجديدة) على المحور الممثل للمتغيرات الجديدة: حيث الإسقاط على المحور Δ_1 والذي يمثل المتغير Z_1 يوضح معامل ارتباط المتغير V_j مع المتغير Z_1.

• دائرة الارتباط بين المتغيرات القديمة V_j والجديدة Z_j

نظرا لأهمية التعرف على طبيعة العلاقة بين المتغيرات القديمة والجديدة يحسب معامل الارتباط بينها. لذا لتمثيل المتغيرات القديمة يؤخذ معامل ارتباطها مع المتغيرات الجديدة. فنحصل على "دائرة الارتباط"، بما أن معامل الارتباط محصور بين -1 و 1 يمكن تمثيل المتغيرات الابتدائية كنقط داخل الدائرة ذات نصف القطر 1.



د\ تحليل العلاقة بين المتغيرات القديمة V_j

د- 1\ نوعية تمثيل المتغيرات:

نقول أن متغير ممثل بطريقة جيدة على محور إذا كان معامل ارتباطه مع المركّب الأساسي بالقيمة المطلقة قريبا (ارتباط قوي). بحيث معامل الارتباط بين المتغير القديم Vc_j والمتغير الجديد Z_k هو جيب تمام الزاوية التي يحددها المحور الممثل عليه والشعاع الرابط بين نقطة المرجع والنقطة v_j التي تمثل المتغير على المحور (أو المستوي أو الفضاء الجزئي) .

نقول أن متغير ممثل جيدا على مستوي إذا كان قريبا من حافة دائرة الارتباط (محيطها). لأن هذا يعني أن القيمة المطلقة لجيب تمام الزاوية بين الشعاع والمستوي تقترب من 1.

د- 2\ دراسة العلاقة بين المتغيرات:

على دائرة الارتباطات يمكن أيضا تفسير مواقع المتغيرات القديمة بالنسبة إلى بعضها في شكل معاملات ارتباط، بشرط أن تكون كل المتغيرات القديمة محل الدراسة ممثلة جيدا على دائرة الارتباط.

- **معامل الارتباط** بين متغيرين قديمين **يساوي** إلى **جيب تمام** الزاوية بين النقطتين الممثلتين لهما على الفضاء الجزئي (1، 2).

$\cos\alpha \to 1 \Rightarrow \alpha \to 0°$.	$\cos\alpha \to -1 \Rightarrow \alpha \to 180°$.	$\cos\alpha \to 0 \Rightarrow \alpha \to 90°$
ارتباط خطي قوي موجب	**ارتباط خطي قوي سالب**	**عدم وجود ارتباط خطي**

ملاحظة هامة: نظرا لكون معامل الارتباط هو مقياس للعلاقة الخطية بين المتغيرات، فإن ضعف معامل الارتباط بين متغيرين لا ينفي وجود ارتباط غير خطي قوي بينهما.

في مثالنا المرفق: القراءة المباشرة لجدول الارتباط بين المتغيرات القديمة والجديدة (ص 32) توضح علاقة ارتباط قوية إيجابية بين المحور وبين كل من التجارة، الإسكان، والزراعة، بينما هناك علاقة ارتباط قوية سلبية بين المحور والدين

العام. يظهر التعاكس بين هذين المجموعتين على دائرة الارتباط المتعلقة بالفضاء الجزئي (1- 2) (ص 35). هذا التعاكس بين DET و {EDU, ACS, AGR, CMI, LOG} وهو الواضح في مصفوفة التباين المشترك (ص 29) يمكن أن تعطي تفسيرا للمحور Δ_1. المحور Δ_2 متعلق كثيرا بالمتغيرEDU، بند الميزانية هذا يظهر تطور مختلف تماما عن البنود الأخرى.

تقارب المتغيرات {EDU, ACS, AGR, CMI, LOG} على دائرة الارتباط (ص 35) يظهر بأن هناك علاقة قوية بين هذه المتغيرات وذلك لأنها ممثلة بطريقة جيدة على الفضاء الجزئي (1- 2) (فالنقط المعنية قريبة من الدائرة).

قرب المتغيرينPVP و TRA من بعضهما لا يمكن تفسيرها لأنها غير ممثلة بطريقة جيدة (فهي بعيدة عن الدائرة). وإن عدنا إلى $\sum$ بينهما (ص 29) سنجده ضعيفا(1.2).

هـ\ تفسير المحاور الأساسية كدالة إلى المتغيرات القديمة:

يمكن تفسير المحاور الأساسية (والتي تمثلها Zj) كدالة إلى المتغيرات القديمة، بحيث يزداد تفسير المتغير القديم Vj لمحور أساسي كلما زاد معامل ارتباطهما R(Z,V).

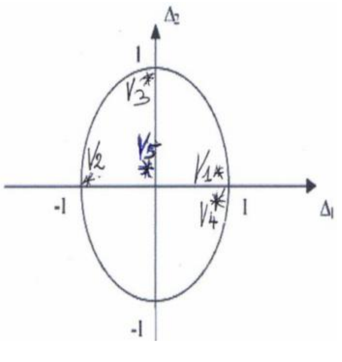
مصفوفة التشبعات أو الشيوعات للمركبة الرئيسية Z_j (أو مصفوفة المركبة في طريقة ACP):

وقيمها عبارة عن معاملات مركبات المتجه المميز Aj المرافقة للجذر المميز الخاص بالمركبة المميزة Zj. والتي هي عبارة عن أوزان أو احتمالات للمعادلة التالية:

$$Z_j = A_{1j}V_{c1} + A_{2j}V_{c2} + \ldots + A_{pj}V_{cp}$$

حيث Aij: مركبات المتجه المميز المرافق للجذر المميز λj الخاص بالمركب Zj.

المصدر: L' analyse des données à l’usage des non mathématiciens (AMVacp)



تمرين: تعطى دائرة الارتباط بين المتغيرات الرئيسية (Zj) والمتغيرات الأولية (القديمة) (Vj):

1- حلل نوعية تمثيل المتغيرات القديمة على الفضاء الجزئي (1،2).
2- ناقش العلاقة بين المتغيرات القديمة فيما بينها.
3- حلل العلاقة بين أول متغيرين قديمين وأول مركبين وأساسيين.
4- أثبت أن: $Cov(Z,V)= \sum A= A \Lambda$.
5- حلل حدود الدراسة (مدى استخدام نتائج الدراسة للتنبؤ).