الفصل الثالث: طريقة التحليل بالمركبات الأساسية

Analyse En Composante Principale (ACP)

تمهيد

نتناول في هذا الفصل طريقة التحليل بالمركبات الأساسية (ACP)، و تستخدم هذه الطريقة في تحليل الجداول الإحصائية بشرط أن تكون كل متغيرات الجدول الإحصائي ذات طبيعة كمية. و تهدف هذه الطريقة إلى البحث عن فضاء شعاعي جزئي اقل درجة عادةً ما يكون ذو البعد 2، يسمح لنا هذا الفضاء الشعاعي الجزئي بأحسن تمثيل ويحفظ لنا اكبر كمية من المعلومات. حيث أننا نستهل هذه الطريقة بتوضيح مبدأ و جوهر الفكرة لطريقة التحليل بالمركبات الأساسية، ثم نتطرق لبعض خصائص و نتائج تحليل الأفراد في الفضاء $\Re^p$، ثم نهتم بدراسة و تحليل خصائص المتغيرات في الفضاء $\Re^n$ و هذا يخص فقط حالة طريقة التحليل بالمركبات الأساسية المرجحة (ACP - Normé).

1. المبدأ الأساسي لطريقة التحليل بالمركبات الأساسية

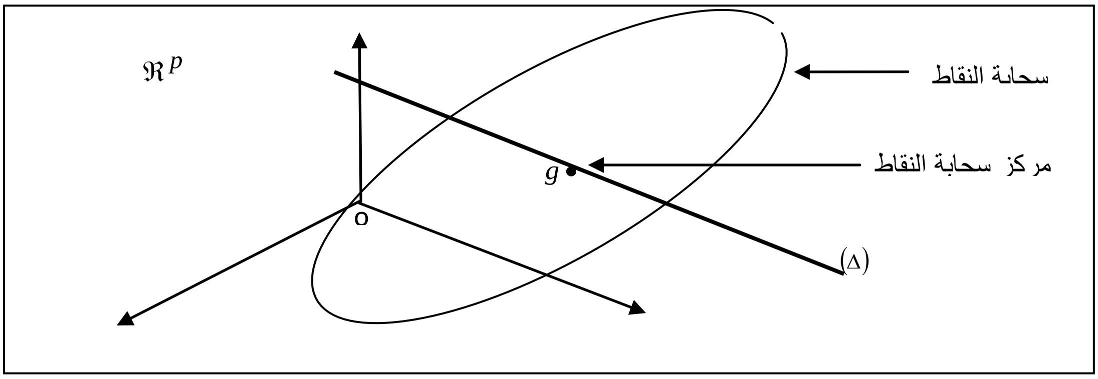
ليكن $R(n,p)$ الجدول الأولي للمعطيات الذي يحتوي على n فرد ممثلين في الأسطر و p متغير أو خاصية ممثلة في الأعمدة، و نكتب:

$$R(n,p)=\begin{array}{c} \\ 1 \\ . \\ . \\ i \\ . \\ n \end{array}\begin{array}{c} \begin{array}{cccccc} 1 & 2 & \ldots\ldots & j & \ldots\ldots & p \end{array} \\ \begin{pmatrix} \eta_{11} & \eta_{12} & .. & \eta_{1j} & .. & \eta_{1p} \\ .. & .. & .. & .. & \ldots & \\ .. & .. & .. & .. & .. & \\ \eta_{i1} & \eta_{i2} & .. & \eta_{ij} & .. & \eta_{ip} \\ .. & .. & .. & .. & & .. \\ \eta_{n1} & \eta_{n2} & .. & \eta_{nj} & .. & \eta_{np} \end{pmatrix}\end{array} \qquad i=\overline{1,n} \quad ; \quad j=\overline{1,p}$$

حيث أن η_{ij} هو العنصر المولد للمصفوفة $R(n,p)$.

كما هو معلوم سابقاً فان المقصود بالتحليل في $\Re^p$ هو دراسة و تحليل الأفراد، فالتمثيل البياني لـ n فرد في الفضاء $\Re^p$ يعطي لنا سحابة نقاط موضحة في الشكل التالي:

الشكل (3) : التمثيل البياني للتحليل في $\Re^p$



و بيانياً المتوسطات الحسابية لـ p متغير تمثل إحداثيات مركز سحابة النقاط g.

$$g = \left(\overline{\eta}_1 \quad \overline{\eta}_2 \quad .... \quad \overline{\eta}_j \quad .... \quad \overline{\eta}_p\right)$$

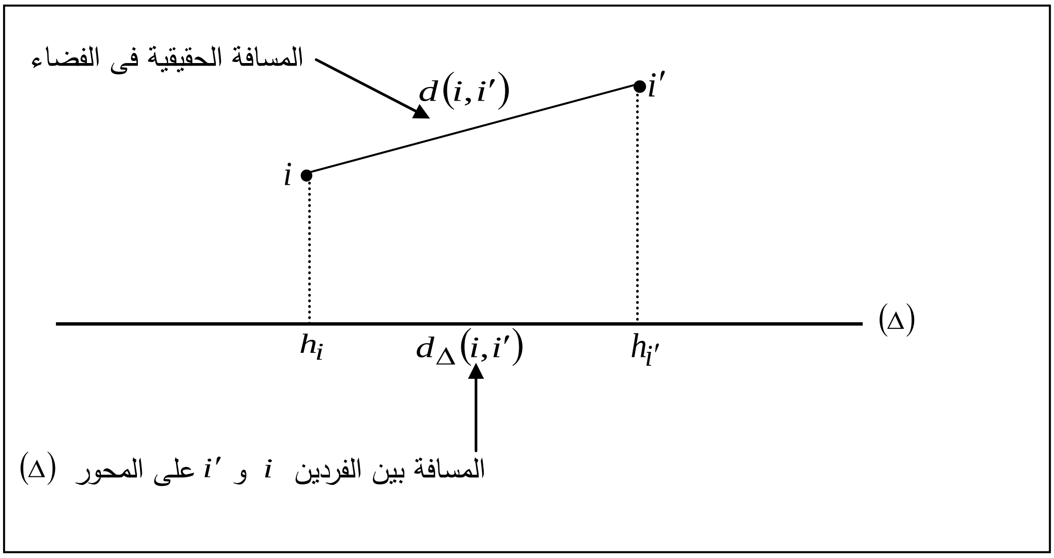
$$\forall j = \overline{1, p} : \ \overline{\eta_j} = \sum_{i=1}^{n} \eta_{ij}$$

إن التمثيل البياني السابق هو صورة توضيحية لـ n فرد في الفضاء $\Re^p$ أي الفضاء الذي يحتوي على p محور، غير أن تحليل و دراسة التمثيل السابق لا يكون إلا على سطح الورقة أو سطح المكتب أي على فضاء شعاعي جزئي ذو البعد 2، و عليه فان الصورة السابقة تصبح غير واضحة المعالم و لا يمكننا ملاحظة أشياء كثيرة منها.

و على أساس هذا التحليل فإننا نبحث عن فضاء شعاعي جزئي اقل درجة عادةً ما يكون ذو البعد 2 (سطح الورقة أو سطح المكتب) يسمح لنا بأحسن تمثيل للبيانات و يحفظ لنا اكبر كمية من المعلومات.

ليكن (Δ) احد محاور الفضاء الشعاعي الجزئي المراد البحث عنه، حيث أن (Δ) يخترق المركز g ، و نريد تعظيم المسافة بين فردين على هذا المحور و التمثيل البياني التالي يوضح هذه الوضعية:

الشكل (4) : التمثيل البياني للإسقاط العمودي على المحور (Δ)



لاحظ أن: $$d^2(i,i') \geq d_\Delta^2(i,i') = (h_i - h_{i'})^2$$

و حتى يكون المحور (Δ) يعطي أحسن تمثيل يجب أن يتم تعظم القيمة:

$$Q = \sum_{i=1}^{n} \sum_{i'=1}^{n} d_\Delta^2(i,i')$$

لدينا:

$$Q=\sum_{i=1}^{n}\sum_{i'=1}^{n}d_{\Delta}^{2}(i,i')=\sum_{i=1}^{n}\sum_{i'=1}^{n}(h_i-h_{i'})^2=\sum_{i=1}^{n}\sum_{i'=1}^{n}\left(h_i^2-2h_ih_{i'}+h_{i'}^2\right)$$

$$=\sum_{i=1}^{n}\sum_{i'=1}^{n}\left(h_i^2\right)-2\sum_{i=1}^{n}\sum_{i'=1}^{n}(h_ih_{i'})+\sum_{i=1}^{n}\sum_{i'=1}^{n}\left(h_{i'}^2\right)$$

$$=n\sum_{i=1}^{n}\left(h_i^2\right)-2n^2\left(\frac{1}{n}\sum_{i=1}^{n}h_i\right)\left(\frac{1}{n}\sum_{i'=1}^{n}h_{i'}\right)+n\sum_{i'=1}^{n}\left(h_{i'}^2\right)$$

و على أساس أن i أو i' يأخذون نفس القيم و هي من 1 إلى n ، يمكننا أن نكتب:

$$\sum_{i=1}^{n}\left(h_i^2\right)=\sum_{i'=1}^{n}\left(h_{i'}^2\right)$$

$$\left(\frac{1}{n}\sum_{i=1}^{n}h_i\right)=\left(\frac{1}{n}\sum_{i'=1}^{n}h_{i'}\right)=\bar{h}$$

حيث أن $\bar{h}$ يمثل إحداثية إسقاط مركز سحابة النقاط g على المحور (Δ).

ومنه يكون:

$$Q=n\sum_{i=1}^{n}\left(h_i^2\right)-2n^2\left(\frac{1}{n}\sum_{i=1}^{n}h_i\right)\left(\frac{1}{n}\sum_{i'=1}^{n}h_{i'}\right)+n\sum_{i'=1}^{n}\left(h_{i'}^2\right)$$

$$=2n\sum_{i=1}^{n}\left(h_i^2\right)-2n^2\bar{h}^2$$

$$=2n\left(\sum_{i=1}^{n}\left(h_i^2\right)-n\bar{h}^2\right)$$

$$=2n\sum_{i=1}^{n}\left(h_i-\bar{h}\right)^2$$

$$=2n\times d_{\Delta}^{2}(i,g)$$

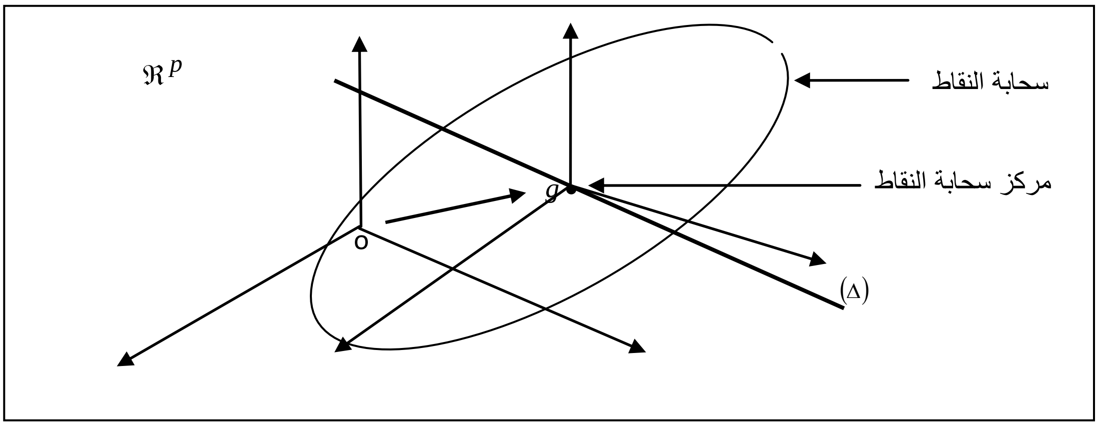
و على اعتبار أن n يمثل عدد الأفراد و هو ثابت يكون تعظيم القيمة $Q=2n\times d_{\Delta}^{2}(i,g)$ يستلزم تعظيم القيمة $d_{\Delta}^{2}(i,g)$ ، و هذه النتيجة تعني تعظيم مسافات الإسقاط بين كل النقاط و مركز السحابة على المحور (Δ).

فإذا قمنا بنقل المعلم من المبدأ (O) إلى مركز سحابة النقاط g ، فان X هي مصفوفة البيانات المركزة و التي تمثل إحداثيات الأفراد في المعلم ذو المبدأ g ، و نكتب:

$$X_{(n,p)}=\begin{array}{c} \\ 1 \\ . \\ . \\ i \\ . \\ n \end{array}\begin{array}{c} \begin{array}{ccccccc} 1 & 2 & .... & j & .... & p \end{array} \\ \begin{pmatrix} x_{11} & x_{12} & .. & x_{1j} & .. & x_{1p} \\ .. & .. & .. & .. & ... & \\ .. & .. & .. & .. & .. & \\ x_{i1} & x_{i2} & .. & x_{ij} & .. & x_{ip} \\ .. & .. & .. & .. & & .. \\ x_{n1} & x_{n2} & .. & x_{nj} & .. & x_{np} \end{pmatrix} \end{array} \qquad ; \quad x_{ij}=\left(\eta_{ij}-\overline{\eta_j}\right)$$

$$i=\overline{1,n} \qquad ; \quad j=\overline{1,p}$$

الشكل (5) : التمثيل البياني لنقل المعلم من المبدأ o إلى مركز سحابة النقاط g



و بعد نقل المعلم من o نحو g يكون $(O\equiv g)$ و يصبح عندئذٍ المشكل المطروح هو تعظيم المسافة بين مبدأ المعلم و إسقاطات النقاط على المحور (Δ) ، و كما تتذكر أن هذا المشكل تمت معالجته في الفصل السابق أي باستعمال طريقة التحليل العاملي العام (AFG).

وعليه فان طريقة التحليل بالمركبات الأساسية (ACP) تقتضي نقل البيانات من المعلم الأصلي ذو المبدأ o نحو المعلم الجديد ذو المبدأ g و تحديد المصفوفة X ثم تطبيق طريقة التحليل العاملي العام (AFG) على المصفوفة X ، غير انه يتم استخراج القيم الذاتية و الأشعة الذاتية من المصفوفة $\left(\frac{X'X}{n}\right)$ وهذا من اجل معاني تفسيرية فقط، و يمكننا أن نميز الحالتين التاليتين:

1.1. طريقة التحليل بالمركبات الأساسية البسيطة (ACP – Non normé)

يمكننا استعمال هذه الطريقة في حالة تجانس المتغيرات، أي أن تكون كل متغيرات الجدول الأولي للمعطيات لها نفس وحدة القياس، وفي هذه الحالة تكون المصفوفة X مصفوفة ممركزة و على النحو التالي:

$$X_{(n,p)} = \begin{array}{c} \\ 1 \\ . \\ . \\ i \\ . \\ n \end{array} \begin{array}{c} \begin{array}{ccccccc} 1 & 2 & \ldots\ldots & j & \ldots\ldots & p \end{array} \\ \begin{pmatrix} x_{11} & x_{12} & .. & x_{1j} & .. & x_{1p} \\ .. & .. & .. & .. & \ldots & \\ .. & .. & .. & .. & .. & \\ x_{i1} & x_{i2} & .. & x_{ij} & .. & x_{ip} \\ .. & .. & .. & .. & & .. \\ x_{n1} & x_{n2} & .. & x_{nj} & .. & x_{np} \end{pmatrix} \end{array} \qquad ; \quad x_{ij} = \left(\eta_{ij} - \overline{\eta_j}\right)$$

$$i = \overline{1,n} \qquad ; \quad j = \overline{1,p}$$

أما استخراج القيم الذاتية و الأشعة الذاتية فيكون من المصفوفة $\left(\frac{X'X}{n}\right)$، ونكتب:

$$\left(\frac{X'X}{n}\right) = \frac{1}{n} \begin{pmatrix} x_{11} & x_{21} & .. & x_{i1} & .. & x_{n1} \\ .. & .. & .. & .. & \ldots & \\ .. & .. & .. & .. & .. & \\ x_{1j} & x_{2j} & .. & x_{ij} & .. & x_{nj} \\ .. & .. & .. & .. & & .. \\ x_{1p} & x_{2p} & .. & x_{ip} & .. & x_{np} \end{pmatrix} \times \begin{pmatrix} x_{11} & x_{12} & .. & x_{1j} & .. & x_{1p} \\ .. & .. & .. & .. & \ldots & .. \\ .. & .. & .. & .. & .. & .. \\ x_{i1} & x_{i2} & .. & x_{ij} & .. & x_{ip} \\ .. & .. & .. & .. & & .. \\ x_{n1} & x_{n2} & .. & x_{nj} & .. & x_{np} \end{pmatrix}$$

$$= \begin{pmatrix} \frac{1}{n}\sum_{i=1}^{n} x_{i1}^2 & \frac{1}{n}\sum_{i=1}^{n} x_{i1}x_{i2} & \ldots\ldots & \frac{1}{n}\sum_{i=1}^{n} x_{i1}x_{ij} & \ldots\ldots & \frac{1}{n}\sum_{i=1}^{n} x_{i1}x_{ip} \\ .. & .. & \ldots\ldots & .. & \ldots\ldots & \\ .. & .. & \ldots\ldots & .. & \ldots\ldots & \\ \frac{1}{n}\sum_{i=1}^{n} x_{i1}x_{ij} & \frac{1}{n}\sum_{i=1}^{n} x_{i2}x_{ij} & \ldots\ldots & \frac{1}{n}\sum_{i=1}^{n} x_{ij}^2 & \ldots\ldots & \frac{1}{n}\sum_{i=1}^{n} x_{ij}x_{ip} \\ .. & .. & \ldots\ldots & .. & \ldots\ldots & .. \\ \frac{1}{n}\sum_{i=1}^{n} x_{i1}x_{ip} & \frac{1}{n}\sum_{i=1}^{n} x_{i2}x_{ip} & \ldots\ldots & \frac{1}{n}\sum_{i=1}^{n} x_{ij}x_{ip} & .. & \frac{1}{n}\sum_{i=1}^{n} x_{ip}^2 \end{pmatrix}$$

إن العنصر المولد للمصفوفة أعلاه هو: $\frac{1}{n}\sum_{i=1}^{n} x_{ij}\,x_{ij'}$

و كما نعلم أن في المصفوفة المركزة فان: $x_{ij} = \left(\eta_{ij} - \overline{\eta_j}\right)$

و بالتعويض يكون العنصر المولد للمصفوفة $\left(\frac{X'X}{n}\right)$ هو: $\frac{1}{n}\sum_{i=1}^{n}\left(\eta_{ij}-\overline{\eta}_j\right)\left(\eta_{ij'}-\overline{\eta}_{j'}\right)$

و القيمة أعلاه تعبر عن التباين المشترك للمتغيرين j و j' ، و نكتب:

$$Cov(j, j') = \frac{1}{n}\sum_{i=1}^{n}\left(\eta_{ij}-\overline{\eta}_j\right)\left(\eta_{ij'}-\overline{\eta}_{j'}\right)$$

و عليه فان المصفوفة $\left(\frac{X'X}{n}\right)$ تعبر عن مصفوفة التباين و التباين المشترك للمتغيرات، و هي مصفوفة مربعة ذات p سطر و p عمود و متناظرة حيث أنها في القطر الأول تحتوي على تباينات المتغيرات أما باقي العناصر فهي التباينات المشتركة للمتغيرات.

و عليه فان مصفوفة التباين و التباين المشترك للمتغيرات تكتب على النحو التالي:

$$V = \frac{X'X}{n} = \begin{array}{c|c} & \begin{array}{cccccccc} 1 & 2 & \cdots\cdots & j & \cdots\cdots & p \end{array} \\ \begin{array}{c} 1 \\ . \\ . \\ j \\ . \\ p \end{array} & \begin{pmatrix} V(1) & Cov(1,2) & .. & Cov(1,j) & .. & Cov(1,p) \\ .. & .. & .. & .. & \cdots & \\ .. & .. & .. & .. & .. & \\ Cov(j,1) & Cov(j,2) & .. & V(j) & .. & Cov(j,p) \\ .. & .. & .. & .. & & .. \\ Cov(p,1) & Cov(p,2) & .. & Cov(p,j) & .. & V(p) \end{pmatrix} \end{array}$$

$j = \overline{1, p}$

2.1. طريقة التحليل بالمركبات الأساسية المرجحة (ACP - Normé)

و يكون من الضروري استعمال هذه الطريقة في حالة عدم تجانس المتغيرات، أي أن تكون متغيرات الجدول الأولي للمعطيات ليس لها نفس وحدة القياس، وفي هذه الحالة تكون المصفوفة X مصفوفة ممركزة ومرجحة و هي على النحو التالي:

$$X_{(n,p)} = \begin{array}{c|c} & \begin{array}{cccccc} 1 & 2 & \cdots & j & \cdots & p \end{array} \\ \begin{array}{c} 1 \\ . \\ . \\ i \\ . \\ n \end{array} & \begin{pmatrix} x_{11} & x_{12} & .. & x_{1j} & .. & x_{1p} \\ .. & .. & .. & .. & \cdots & \\ .. & .. & .. & .. & .. & \\ x_{i1} & x_{i2} & .. & x_{ij} & .. & x_{ip} \\ .. & .. & .. & .. & & .. \\ x_{n1} & x_{n2} & .. & x_{nj} & .. & x_{np} \end{pmatrix} \end{array} \quad ; \quad x_{ij} = \frac{\left(\eta_{ij}-\overline{\eta}_j\right)}{\sigma_j}$$

$i = \overline{1, n} \quad ; \quad j = \overline{1, p}$

حيث أن σ_j يعبر عن الانحراف المعياري للمتغير j .

و يتم كذلك بنفس الطريقة استخراج القيم الذاتية و الأشعة الذاتية من المصفوفة $\left(\frac{X'X}{n}\right)$، وبنفس خطوات حساب $\left(\frac{X'X}{n}\right)$ يتم تحديد عناصر هذه المصفوفة غير أنها في هذه الحالة تكون على النحو التالي:

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\left(\eta_{ij}-\overline{\eta}_j\right)\left(\eta_{ij'}-\overline{\eta}_{j'}\right)}{\sigma_j\times\sigma_{j'}}$$

و القيمة أعلاه تعبر عن معامل الارتباط الخطي البسيط بين المتغيرين j و j' ، و نكتب:

$$r(j,j')=\frac{1}{n}\sum_{i=1}^{n}\frac{\left(\eta_{ij}-\overline{\eta}_j\right)\left(\eta_{ij'}-\overline{\eta}_{j'}\right)}{\sigma_j\times\sigma_{j'}}$$

و عليه فان المصفوفة $\left(\frac{X'X}{n}\right)$ تعبر عن مصفوفة الارتباطات، و هي مصفوفة مربعة ذات p سطر و p عمود و متناظرة حيث أنها في القطر الأول تحتوي على القيمة واحد و هي تعني معامل الارتباط الخطي البسيط للمتغير مع نفسه أما باقي العناصر فهي معاملات الارتباط الخطي البسيط للمتغيرات فيما بينها، و نكتب:

$$C=\frac{X'X}{n}=\begin{array}{c} \\ 1\\ .\\ .\\ j\\ .\\ p\end{array}\begin{array}{c}\begin{array}{cccccc}1 & 2 & \ldots\ldots & j & \ldots\ldots\ldots\ldots & p\end{array}\\ \begin{pmatrix}1 & r(1,2) & .. & r(1,j) & .. & r(1,p)\\ .. & .. & .. & .. & \ldots & \\ .. & .. & .. & .. & .. & \\ r(j,1) & r(j,2) & .. & 1 & .. & r(j,p)\\ .. & .. & .. & .. & & ..\\ r(p,1) & r(p,2) & .. & r(p,j) & .. & 1\end{pmatrix}\end{array}$$

$$j=\overline{1,p}$$

2. التحليل في $\Re^p$

للتحليل في $\Re^p$ نهتم بالخصائص التالية للمركبات الأساسية:

1.2. المتوسط الحسابي للمركبات الأساسية

تعطى عبارة المركبات الأساسية للأفراد على المحور α على النحو التالي:

$$F_\alpha=XU_\alpha=\left(X^1\ldots\ldots\ldots\ldots\quad X^j\ldots\ldots\ldots\ldots\quad X^p\right)\times\begin{pmatrix}U_\alpha(1)\\ .\\ .\\ U_\alpha(j)\\ .\\ .\\ U_\alpha(p)\end{pmatrix}=\sum_{j=1}^{p}X^jU_\alpha(j)$$

و يكون المتوسط الحسابي للمركبات الأساسية هو:

$$\forall \alpha = \overline{1,p}: \quad \overline{F_\alpha} = \sum_{j=1}^{p} \overline{X}_j U_\alpha(j) = 0$$

و ذلك لان $X(n,p)$ مصفوفة ممركزة: $\overline{X}_j = 0$

2.2. تباين المركبات الأساسية

$$Var(F_\alpha) = \frac{1}{n}\sum_{\alpha=1}^{p}\left(F_\alpha - \overline{F}_\alpha\right)^2 = \frac{1}{n}\sum_{\alpha=1}^{p}(F_\alpha)^2 = \frac{1}{n}F'_\alpha F_\alpha = \frac{1}{n}(XU_\alpha)'(XU_\alpha)$$

$$= U_\alpha'\left(\frac{X'X}{n}\right)U_\alpha \qquad \Big/ \qquad \left(\frac{X'X}{n}\right)U_\alpha = \lambda_\alpha U_\alpha$$

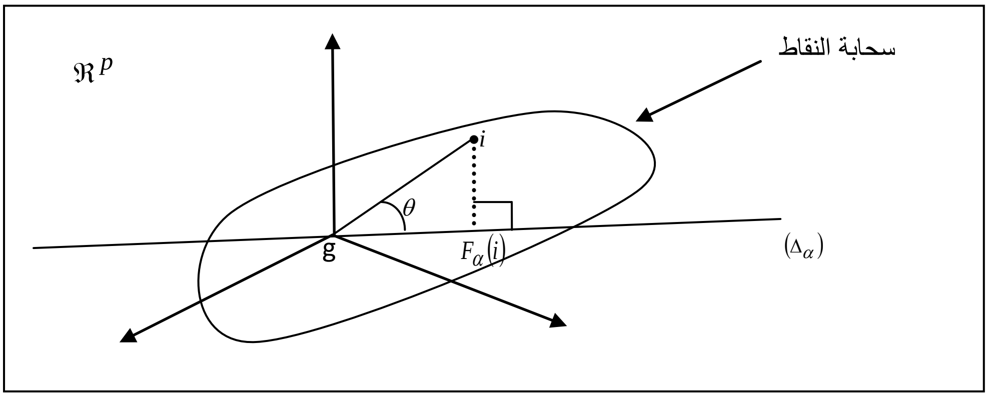
$$= \lambda_\alpha\left(U_\alpha' U_\alpha\right) = \lambda_\alpha$$

أي أن تباين المركبة الأساسية يساوي القيمة الذاتية للمحور.

3.2. نسب تمثيل الأفراد على المحاور

نعلم مسبقاً انه كلما كانت القيمة $|F_\alpha(i)|$ كبيرة كان المحور (Δ_α) جيد في تمثيل الفرد i، وهذا يجعل من النقطة i أكثر قرباً من المحور (Δ_α) وعندئذٍ تتقلص الزاوية θ، و الشكل البياني يوضح ذلك:

الشكل (6) : التمثيل البياني لنسب تمثيل الأفراد على المحاور



و عليه فإننا نقترح $Cos^2\theta_{i\alpha}$ كمؤشر يقيس نسبة تمثيل الفرد i على المحور α، و نكتب:

$$Cos^2\theta_{i\alpha} = \frac{F_\alpha^2(i)}{d^2(g,i)} \qquad ; \qquad d^2(g,i) = \sum_{\alpha=1}^{p} F_\alpha^2(i)$$

و كما نعلم أن:

$$-1 \le Cos\theta_{i\alpha} \le 1 \Rightarrow 0 \le Cos^2\theta_{i\alpha} \le 1$$

و يمكننا القول انه كلما كانت الزاوية $\theta_{i\alpha}$ صغيرة كانت نسبة تمثيل الفرد i على المحور α جيدة و هذا يجعل من $Cos^2\theta_{i\alpha}$ يقترب من الواحد، أما في الحالة العكسية عندما تكون الزاوية $\theta_{i\alpha}$ كبيرة تكون نسبة تمثيل الفرد i على المحور α سيئة و هذا يجعل من $Cos^2\theta_{i\alpha}$ يقترب من الصفر.

أما عن نسب تمثيل الفرد i على المستوي الأول فهي:

$$Cos^2\theta_{i1} + Cos^2\theta_{i2}$$

و على أساس أن مجموع نسب تمثيل الفرد i على المحور α هو 100% يكون: $\sum_{\alpha=1}^{p} Cos^2\theta_{i\alpha} = 1$

4.2. نسب مساهمة الأفراد في تشكيل المحاور

نقيس نسبة مساهمة الفرد i في تشكيل المحور α بالعلاقة التالية:

$$C_i^{\alpha} = \frac{F_{\alpha}^2(i)}{\sum_{i=1}^{n} F_{\alpha}^2(i)} = \frac{F_{\alpha}^2(i)}{F'_{\alpha}F_{\alpha}} = \frac{F_{\alpha}^2(i)}{n\lambda_{\alpha}}$$

$C_i^{\alpha}\%$: نسبة مساهمة الفرد i في تشكيل المحور α

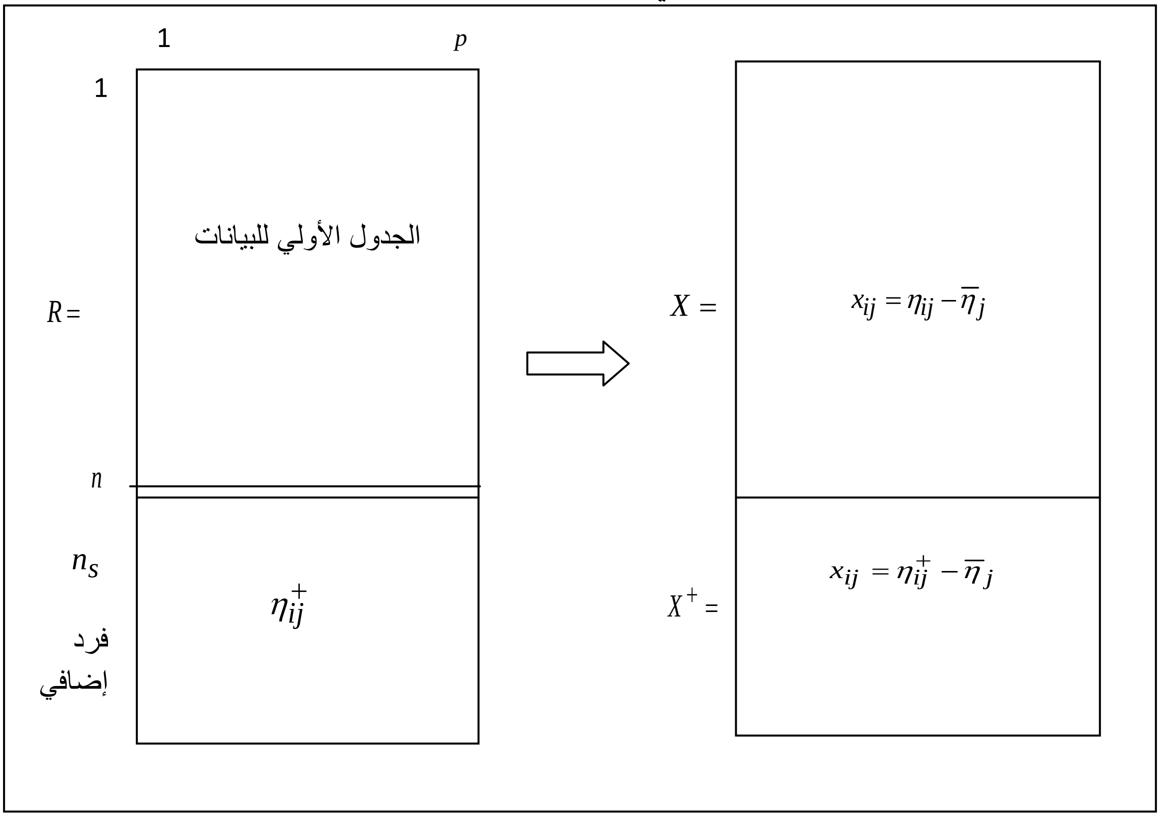
و على أساس أن مجموع نسب مساهمة كل الأفراد في تشكيل المحور α هي 100% يكون:

$$\sum_{i=1}^{n} C_i^{\alpha} = \frac{\sum_{i=1}^{n} F_{\alpha}^2(i)}{\sum_{i=1}^{n} F_{\alpha}^2(i)} = 1$$

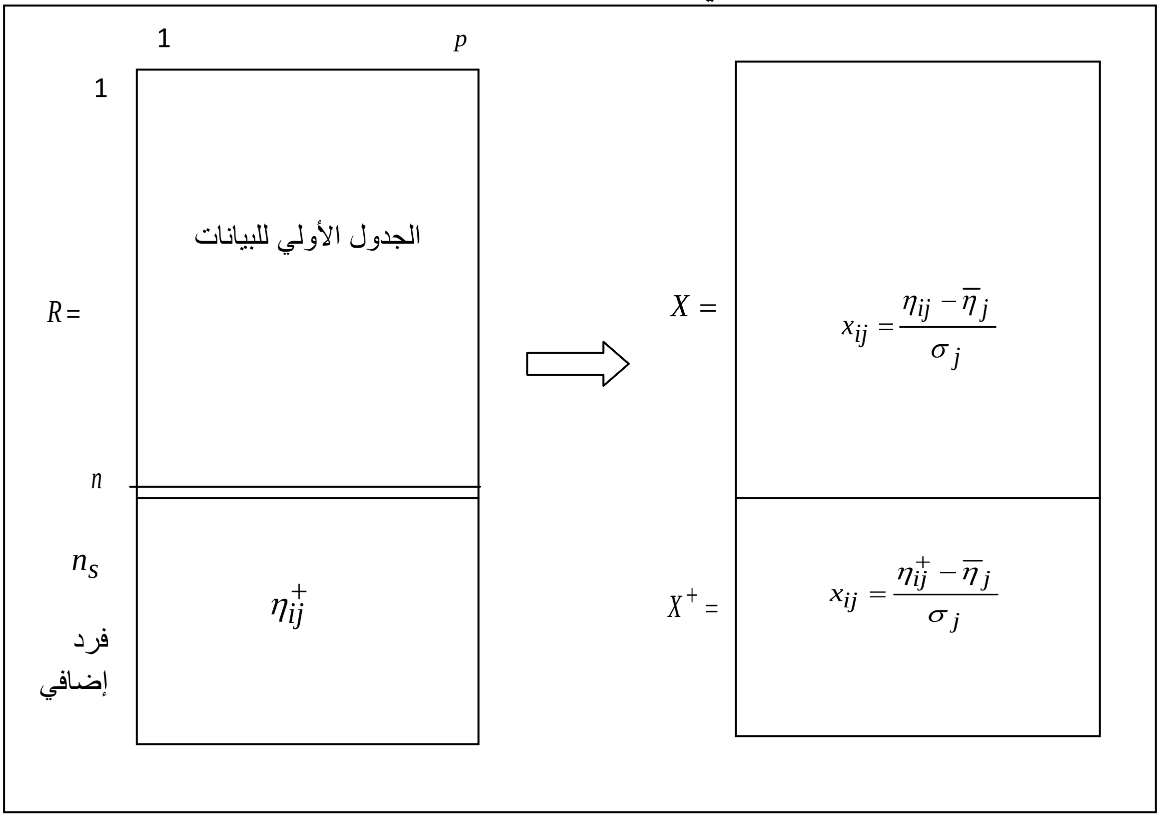
5.2. حالة إضافة أفراد جُدُدْ

في هذا الفرع نعمل على توضيح كيفية حساب إحداثيات الأفراد الجدد وذلك بالاعتماد على نتائج تحليل الجدول الأصلي للمعطيات، و الشكل التالي يوضح ذلك:

الشكل (7) : حالة طريقة التحليل بالمركبات الأساسية البسيطة



الشكل (8) : حالة طريقة التحليل بالمركبات الأساسية المرجحة



و تعطى عبارة المركبات الأساسية للأفراد الجدد على المحاور على النحو التالية:

$$\underset{n_s \times 1}{F_\alpha^+} = \underset{n_s \times p}{X^+} \times \underset{p \times 1}{U_\alpha}$$

3. التحليل في $\Re^n$

نهتم في هذا الفرع بطريقة التحليل بالمركبات الأساسية المرجحة (ACP - Normé) فقط، وعندئذٍ يتم استخراج القيم الذاتية و الأشعة الذاتية من مصفوفة الارتباطات C، أما المصفوفة X المركزة و المرجحة تكون على النحو التالي:

$$X_{(n,p)} = \begin{array}{c} \\ 1 \\ . \\ . \\ i \\ . \\ n \end{array} \begin{array}{c} \begin{array}{cccccc} 1 & 2 & \ldots\ldots & j & \ldots\ldots & p \end{array} \\ \begin{pmatrix} x_{11} & x_{12} & .. & x_{1j} & .. & x_{1p} \\ .. & .. & .. & .. & \ldots & \\ .. & .. & .. & .. & .. & \\ x_{i1} & x_{i2} & .. & x_{ij} & .. & x_{ip} \\ .. & .. & .. & .. & & .. \\ x_{n1} & x_{n2} & .. & x_{nj} & .. & x_{np} \end{pmatrix} \end{array} \quad ; \quad x_{ij} = \frac{\left(\eta_{ij} - \overline{\eta_j}\right)}{\sigma_j}$$

$$i = \overline{1, n} \quad ; \quad j = \overline{1, p}$$

حيث أن σ_j يعبر عن الانحراف المعياري للمتغير j .

غير أننا بغيت تحقيق بعض النتائج نعمد إلى القيام بتحويل بسيط على المصفوفة X ، وهذا التحويل لا يخص إلا هذا الفرع، إذن لنضع:

$$C = \frac{X'X}{n} = \left(\frac{X'}{\sqrt{n}}\right)\left(\frac{X}{\sqrt{n}}\right) = \tilde{X}'\tilde{X}$$

حيث أن:

$$\begin{array}{c} \\ \\ \tilde{X}_{(n,p)} = \end{array} \begin{array}{c|cccccc} & 1 & 2 & \ldots\ldots & j & \ldots\ldots & p \\ \hline 1 & \tilde{x}_{11} & \tilde{x}_{12} & .. & \tilde{x}_{1j} & .. & \tilde{x}_{1p} \\ . & .. & .. & .. & .. & \cdots & \\ . & .. & .. & .. & .. & .. & \\ i & \tilde{x}_{i1} & \tilde{x}_{i2} & .. & \tilde{x}_{ij} & .. & \tilde{x}_{ip} \\ . & .. & .. & .. & .. & & .. \\ n & \tilde{x}_{n1} & \tilde{x}_{n2} & .. & \tilde{x}_{nj} & .. & \tilde{x}_{np} \end{array} \quad ; \quad \tilde{x}_{ij} = \frac{\left(\eta_{ij} - \overline{\eta_j}\right)}{\sigma_j \times \sqrt{n}}$$

$$i = \overline{1,n} \quad ; \quad j = \overline{1,p}$$

1.3. إحداثيات المتغيرات على المحاور

نعمل في هذا الفرع على تحديد إحداثيات المتغيرات على المحور α و نتطرق لدلالة هذه الإحداثيات، و نعلم أن عبارة الإحداثيات على النحو التالي:

$$G_\alpha(j) = \tilde{X}'V_\alpha$$

و نعلم أن:

$$V_\alpha = \frac{1}{\sqrt{\lambda_\alpha}}\tilde{X}U_\alpha$$

و بتعويض نجد أن:

$$G_\alpha(j) = \tilde{X}'\frac{1}{\sqrt{\lambda_\alpha}}\tilde{X}U_\alpha$$

و مع أن:

$$\tilde{X} = \frac{1}{\sqrt{n}}X \qquad \text{و} \qquad F_\alpha = XU_\alpha$$

و بتعويض نجد أن:

$$G_\alpha(j) = \frac{1}{\sqrt{\lambda_\alpha}} \times \frac{1}{\sqrt{n}}\tilde{X}'F_\alpha = \frac{1}{\sqrt{\lambda_\alpha}} \times \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\tilde{x}_{ij}F_\alpha(i)$$

و مع العلم أن:

$$\tilde{x}_{ij} = \frac{\eta_{ij} - \overline{\eta}_j}{\sigma_j\sqrt{n}} \qquad \text{و} \qquad \overline{F}_\alpha = 0 \qquad V(F_\alpha) = \lambda_\alpha$$

و بتعويض نجد أن:

$$G_\alpha(j) = \frac{1}{\sqrt{\lambda_\alpha}} \times \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \tilde{x}_{ij} F_\alpha(i)$$

$$= \frac{1}{\sqrt{\lambda_\alpha}} \times \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(\frac{\eta_{ij} - \overline{\eta}_j}{\sigma_j \sqrt{n}} \right) \left(F_\alpha(i) - \overline{F}_\alpha \right)$$

$$= \frac{\frac{1}{n} \sum_{i=1}^{n} \left(\eta_{ij} - \overline{\eta}_j \right) \left(F_\alpha(i) - \overline{F}_\alpha \right)}{\sigma_j \times \sigma(F_\alpha)} = r(\alpha, j)$$

و هذا يعني إن إحداثية المتغير j على المحور α $G_\alpha(j)$ تعبر عن معامل الارتباط الخطي البسيط بين المتغير j و المحور α أي أن $G_\alpha(j) = r(\alpha, j)$.

2.3. المسافة بين المبدأ و المتغيرات

نقوم في هذا الفرع بحساب المسافة بين المبدأ $(g \equiv o)$ و المتغيرات، أي نحسب المسافة $d(g, j)$، و نكتب:

لدينا: إحداثيات المبدأ هي: $g(0 \ldots\ldots\ldots\ 0 \ldots\ldots\ldots\ 0)$

إحداثيات المتغير j هي: $j(\tilde{x}_{1j} \ldots\ldots\ldots\ \tilde{x}_{ij} \ldots\ldots\ldots\ \tilde{x}_{nj})$

تكون المسافة على النحو التالي:

$$d(g, j) = \sqrt{\sum_{i=1}^{n} \left(\tilde{x}_{ij} - 0 \right)^2} = \sqrt{\sum_{i=1}^{n} \left(\frac{\eta_{ij} - \overline{\eta}_j}{\sigma_j \sqrt{n}} \right)^2} = \sqrt{\frac{\frac{1}{n} \sum_{i=1}^{n} \left(\eta_{ij} - \overline{\eta}_j \right)^2}{\sigma_j^2}}$$

و مع العلم أن:

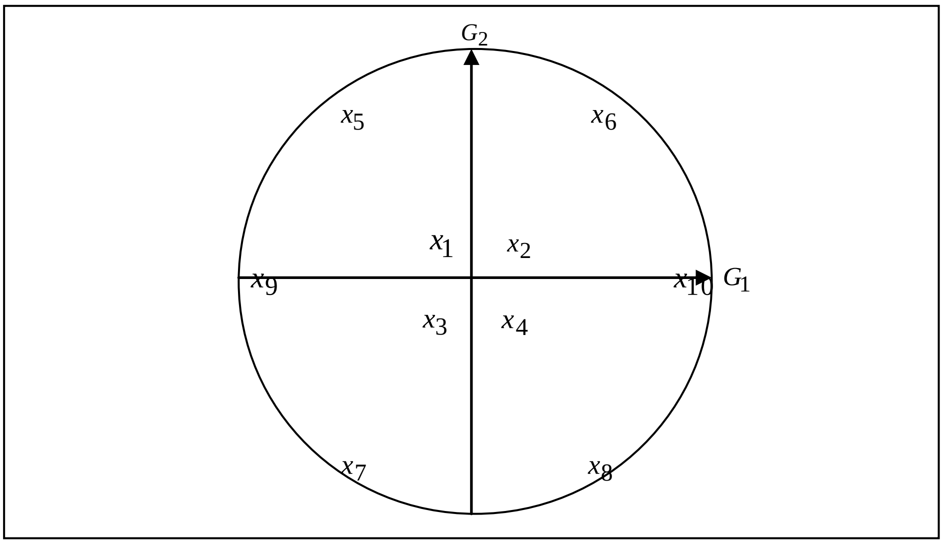
$$\sigma_j^2 = \frac{1}{n} \sum_{i=1}^{n} \left(\eta_{ij} - \overline{\eta}_j \right)^2$$

و بتعويض نجد أن:

$$d(g, j) = \sqrt{\frac{\sigma_j^2}{\sigma_j^2}} = 1$$

وهذه النتيجة تعني أن كل المتغيرات تقع على سطح كرة مركزها المبدأ $(g \equiv o)$ و نصف قطرها يساوي الواحد. و إسقاط هذه الكرة على المستوي الأول و هو المستوي الأفضل على الإطلاق، يعطينا قرص مركزها المبدأ $(g \equiv o)$ و نصف قطرها يساوي الواحد، و الشكل التالي يوضح ذلك:

الشكل (9) : التمثيل البياني لإسقاط المتغيرات على المستوي



و على أساس النتيجة السابقة فان كل المتغيرات تقع داخل هذا القرص و من غير الممكن أن تقع خارجه، فالمتغيرات (x_1, x_2, x_3, x_4) و التي تقع بالقرب من المركز سيئة التمثيل على المستوي المقترح للدراسة و سبب ذلك أن هذه المتغيرات في وضعها الحقيقي تقع في اعلي أو أسفل الكرة أي أنها بعيدة جداً عن المستوي الأول و بالتالي تستثنى من التحليل و الدراسة، أما باقي المتغيرات و التي تقع بالقرب من محيط القرص $(x_5, x_6, x_7, x_8, x_9, x_{10})$ جيدة التمثيل على المستوي المقترح للدراسة و سبب ذلك أن هذه المتغيرات في وضعها الحقيقي تقع بالقرب من المستوي الأول و بالتالي مقبولة في التحليل و الدراسة.

3.3. المسافة بين متغيرين

نقوم في هذا الفرع بحساب المسافة بين متغيرين، أي نحسب المسافة $d(j, j')$ ، و نكتب:

لدينا: إحداثيات المتغير j هي: $j(\tilde{x}_{1j} \ldots\ldots\ldots \quad \tilde{x}_{ij} \ldots\ldots\ldots \quad \tilde{x}_{nj})$

إحداثيات المتغير j' هي: $j'(\tilde{x}_{1j'} \ldots\ldots\ldots \quad \tilde{x}_{ij'} \ldots\ldots\ldots \quad \tilde{x}_{nj'})$

تكون المسافة على النحو التالي:

$$d(j, j') = \sqrt{\sum_{i=1}^{n} (\tilde{x}_{ij} - \tilde{x}_{ij'})^2} = \sqrt{\sum_{i=1}^{n} \tilde{x}_{ij}^2 - 2\sum_{i=1}^{n} \tilde{x}_{ij}\tilde{x}_{ij'} + \sum_{i=1}^{n} \tilde{x}_{ij'}^2}$$

و مع العلم أن:

$$\sum_{i=1}^{n} \tilde{x}_{ij}^2 = \sum_{i=1}^{n} \tilde{x}_{ij'}^2 = 1 \quad \text{و} \quad \tilde{x}_{ij} = \frac{\eta_{ij} - \overline{\eta}_j}{\sigma_j \sqrt{n}} \quad \text{و} \quad \tilde{x}_{ij'} = \frac{\eta_{ij'} - \overline{\eta}_{j'}}{\sigma_{j'} \sqrt{n}}$$

و بتعويض نجد أن:

$$d(j, j') = \sqrt{\sum_{i=1}^{n} \left(\tilde{x}_{ij} - \tilde{x}_{ij'}\right)^2} = \sqrt{2 - 2\sum_{i=1}^{n} \left(\frac{\eta_{ij} - \overline{\eta}_j}{\sigma_j \sqrt{n}}\right) \times \left(\frac{\eta_{ij'} - \overline{\eta}_{j'}}{\sigma_{j'} \sqrt{n}}\right)}$$

$$= \sqrt{2 - 2\frac{\frac{1}{n}\sum_{i=1}^{n} \left(\eta_{ij} - \overline{\eta}_j\right) \times \left(\eta_{ij'} - \overline{\eta}_{j'}\right)}{\sigma_j \times \sigma_{j'}}}$$

$$d(j, j') = \sqrt{2\left(1 - r_{j,j'}\right)}$$

حيث أن $r_{j,j'}$ يمثل معامل الارتباط الخطي البسيط بين المتغيرين j و j' .

$$-1 \le r_{j,j'} \le 1 \Rightarrow 0 \le d(j, j') \le 2$$

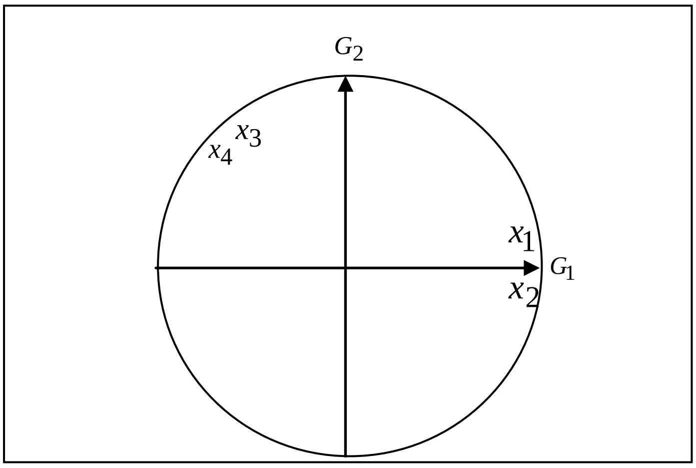
و هذه النتيجة تعني أن المسافة بين المتغيرات تعطي بدلالة الارتباط فيما بينهم، و على أساس هذه النتيجة يمكننا أن نميز الحالات الثلاثة التالية:

A. حالة الارتباط القوي الموجب

إذا كانت المسافة بين المتغيرين j و j' تؤول إلى الصفر فان الارتباط بين هاذين المتغيرين قوي و موجب، و نكتب:

$$\left[d(j, j') \to 0\right] \Leftrightarrow \left[r_{(j,j')} \to 1\right]$$

الشكل (10) : حالة الارتباط القوي الموجب



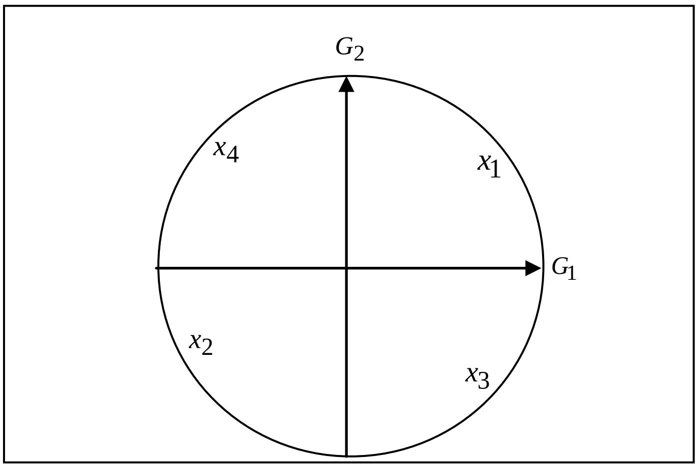
و في الشكل أعلاه يمكننا القول أن المتغيرين x_1 و x_2 يرتبطان ارتباط قوي و موجب، و نفس الشيء بالنسبة للمتغيرين x_3 و x_4 لهما ارتباط قوي و موجب فيما بينهما.

B. حالة الارتباط القوي السالب

إذا كانت المسافة بين المتغيرين j و j' أعظمية فان الارتباط بين هاذين المتغيرين قوي و سالب، و نكتب:

$$\left[d(j, j') \to 2\right] \Leftrightarrow \left[r(j, j') \to -1\right]$$

الشكل (11) : حالة الارتباط القوي السالب



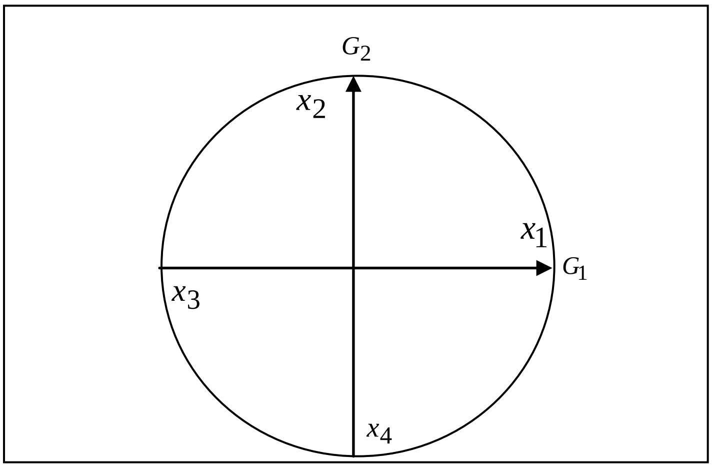
و في الشكل أعلاه يمكننا القول أن المتغيرين x_1 و x_2 يرتبطان ارتباط قوي و سالب، و نفس الشيء بالنسبة للمتغيرين x_3 و x_4 لهما ارتباط قوي و سالب فيما بينهما.

C. حالة الاستقلالية

إذا كانت المسافة بين المتغيرين j و j' تؤول إلى العدد $\sqrt{2}$ فان هاذين المتغيرين مستقلين عن بعض، و نكتب:

$$\left[d(j, j') \to \sqrt{2}\right] \Leftrightarrow \left[r(j, j') \to 0\right]$$

الشكل (12) : حالة الاستقلالية

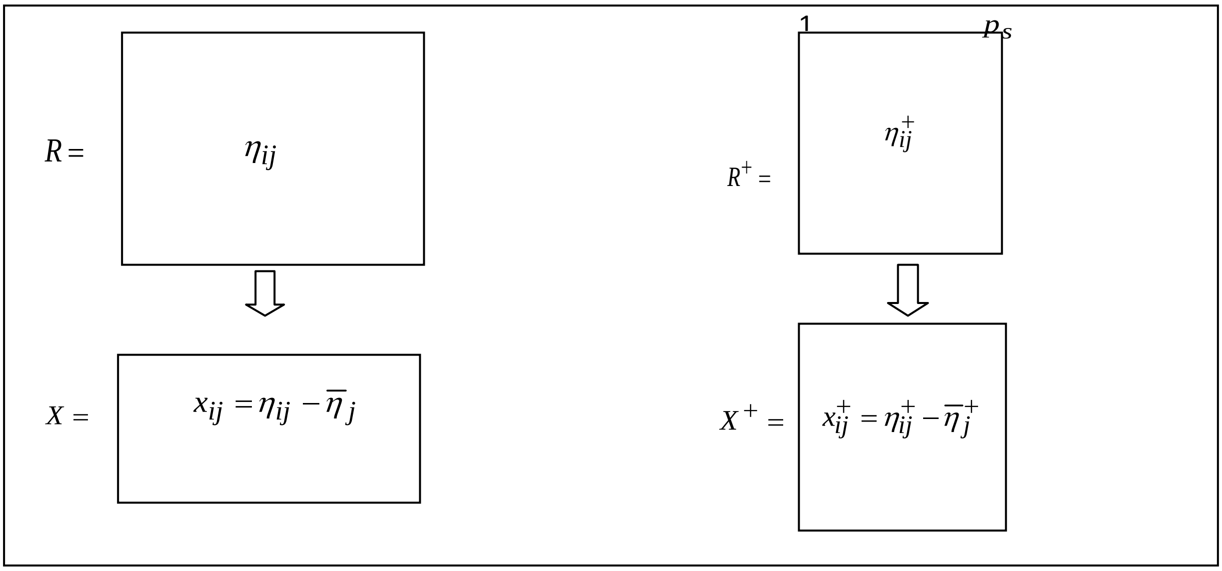


و في الشكل أعلاه يمكننا القول أن المتغيرين x_1 و x_2 مستقلين عن بعض، و نفس الشيء بالنسبة للمتغيرين x_3 و x_4 فهما مستقلين عن بعض.

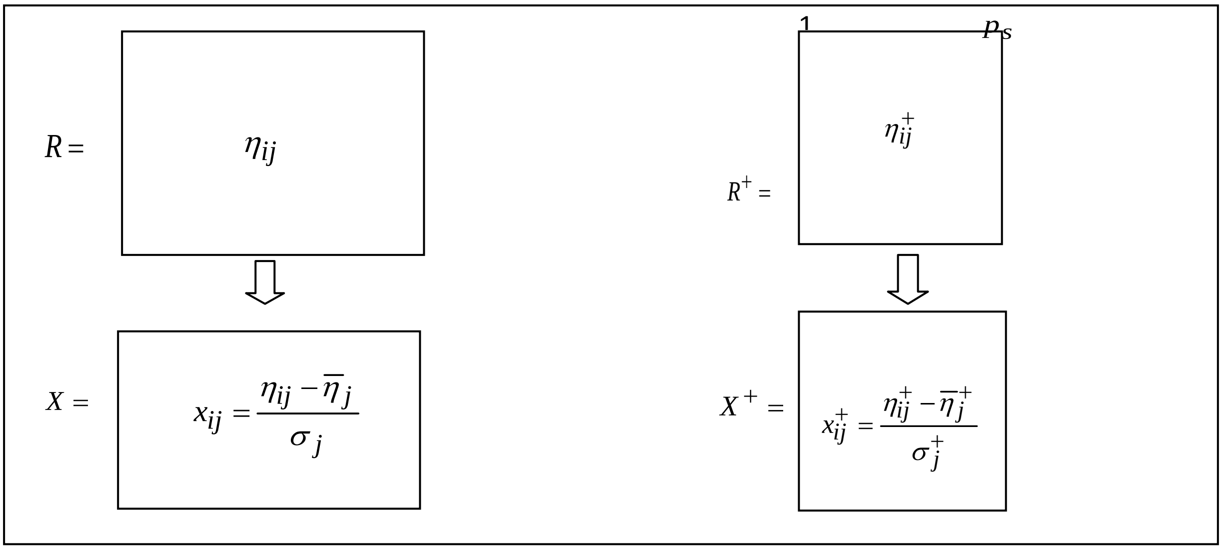
4. حالة إضافة متغيرات جُدُدْ

في هذا الفرع نعمل على توضيح كيفية حساب إحداثيات المتغيرات الجدد وذلك بالاعتماد على نتائج تحليل الجدول الأصلي للمعطيات، و الشكل التالي يوضح ذلك:

الشكل (13) : حالة طريقة التحليل بالمركبات الأساسية البسيطة



الشكل (14) : حالة طريقة التحليل بالمركبات الأساسية المرجحة



و تعطى عبارة إحداثيات المتغيرات الجدد على المحاور على النحو التالية:

$$\underset{p_s \times 1}{G_\alpha^+} = \underset{p_s \times n}{X^{+'}} \times \underset{n \times 1}{V_\alpha}$$

و الملخص التالي يوضح كل خطوات طريقة التحليل بالمركبات الأساسية

الجدول الأولي للمعطيات $R_{(n.p)}$:

$X = \lfloor \eta_{ij} - \overline{\eta}_j \rfloor$ نستعمل هذه المصفوفة في حالة **(ACP- Non normé)**

$X = \left[\dfrac{\eta_{ij} - \overline{\eta}_j}{s_j} \right]$ نستعمل هذه المصفوفة في حالة **(ACP- Normé)** هذه الطريقة ضرورية في حالة عدم تجانس المتغيرات

التحليل في الفضاء الشعاعي R^p:

تحليل الأفراد في فضاء المتغيرات

$$\left(\frac{1}{n} X'X\right) U_\alpha = \lambda_\alpha U_\alpha$$

$U'_\alpha U_\alpha = 1$; *Inertie totale (It) :*

$$\sum_{\alpha=1}^{p} \lambda_\alpha = Trace\left(\frac{1}{n} X'X\right)$$

λ_α : القيم الذاتية للمصفوفة $\frac{1}{n} X'X$

U_α : الأشعة الذاتية المرفقة بالقيمة الذاتية λ_α للمصفوفة $\frac{1}{n} X'X$

نأخذ قيم λ_α بالترتيب التنازلي

$F_\alpha = X\, U_\alpha$: المركبات الأساسية لكل الأفراد على المحور α

$\overline{F}_\alpha = 0$ ، $V(F_\alpha) = \lambda_\alpha$

$\tau_\alpha = \lambda_\alpha \Big/ \sum_{\alpha=1}^{p} \lambda_\alpha = \lambda_\alpha / It$: نسبة تمثيل البيانات على المحور α

$Cos^2\theta_{i\alpha} = F^2{}_\alpha(i) \big/ d^2(i,g)$ نسبة تمثيل الفرد i على المحور α

$d^2(i,g) = \sum_{\alpha=1}^{p} F^2{}_\alpha(i)$ ، $\sum_{\alpha=1}^{p} Cos^2\theta_{i\alpha} = 1$

$C_i^\alpha = F^2{}_\alpha(i) \big/ n\lambda_\alpha$

التحليل في الفضاء الشعاعي R^n: تحليل المتغيرات في فضاء الأفراد

$$\left(\frac{1}{n} XX'\right) V_\alpha = \lambda_\alpha V_\alpha$$

$V'_\alpha V_\alpha = 1$

λ_α : القيم الذاتية للمصفوفة $\frac{1}{n} XX'$

V_α : الأشعة الذاتية المرفقة بالقيمة الذاتية λ_α للمصفوفة $\frac{1}{n} XX'$

$G_\alpha = X' V_\alpha$: المركبات الأساسية لكل المتغيرات على المحور α

$G_\alpha(j) = Cor(j, F_\alpha)$: تمثل معامل الارتباط بين التغير j و المحور α

$d(j,o) = 1$: المسافة بين أي متغير و المركز تساوى 1

$d^2(j,j') = 2.(1 - r_{j,j'})$: المسافة مربعة بين متغيرين j, j'

نهتم فقط بالمتغيرات التي تقع بالقرب من خط الدائرة

$d\,(j,j') = \sqrt{2}$: j مستقل عن j'

$d\,(j,j') = 0$: j يرتبط ارتباط قوي وموجب بـ j'

$d\,(j,j') = 2$: j يرتبط ارتباط قوي وسالب بـ j'

عبارة إعادة تشكيل البيانات:

$$X = \sum_{\alpha=1}^{p} \sqrt{\lambda_\alpha}\, V_\alpha U'_\alpha = \sum_{\alpha=1}^{1} \frac{1}{\sqrt{\lambda_\alpha}} F_\alpha G'_\alpha$$

عبارات أول انتقال

$$V_\alpha = \frac{1}{\sqrt{\lambda_\alpha}} X U_\alpha = \frac{1}{\sqrt{\lambda_\alpha}} F_\alpha \quad , \quad F_\alpha = \sqrt{\lambda_\alpha}\, V_\alpha$$

عبارات ثاني انتقال

$$U_\alpha = \frac{1}{\sqrt{\lambda_\alpha}} X'V_\alpha = \frac{1}{\sqrt{\lambda_\alpha}} G_\alpha \quad , \quad G_\alpha = \sqrt{\lambda_\alpha}\, U_\alpha$$