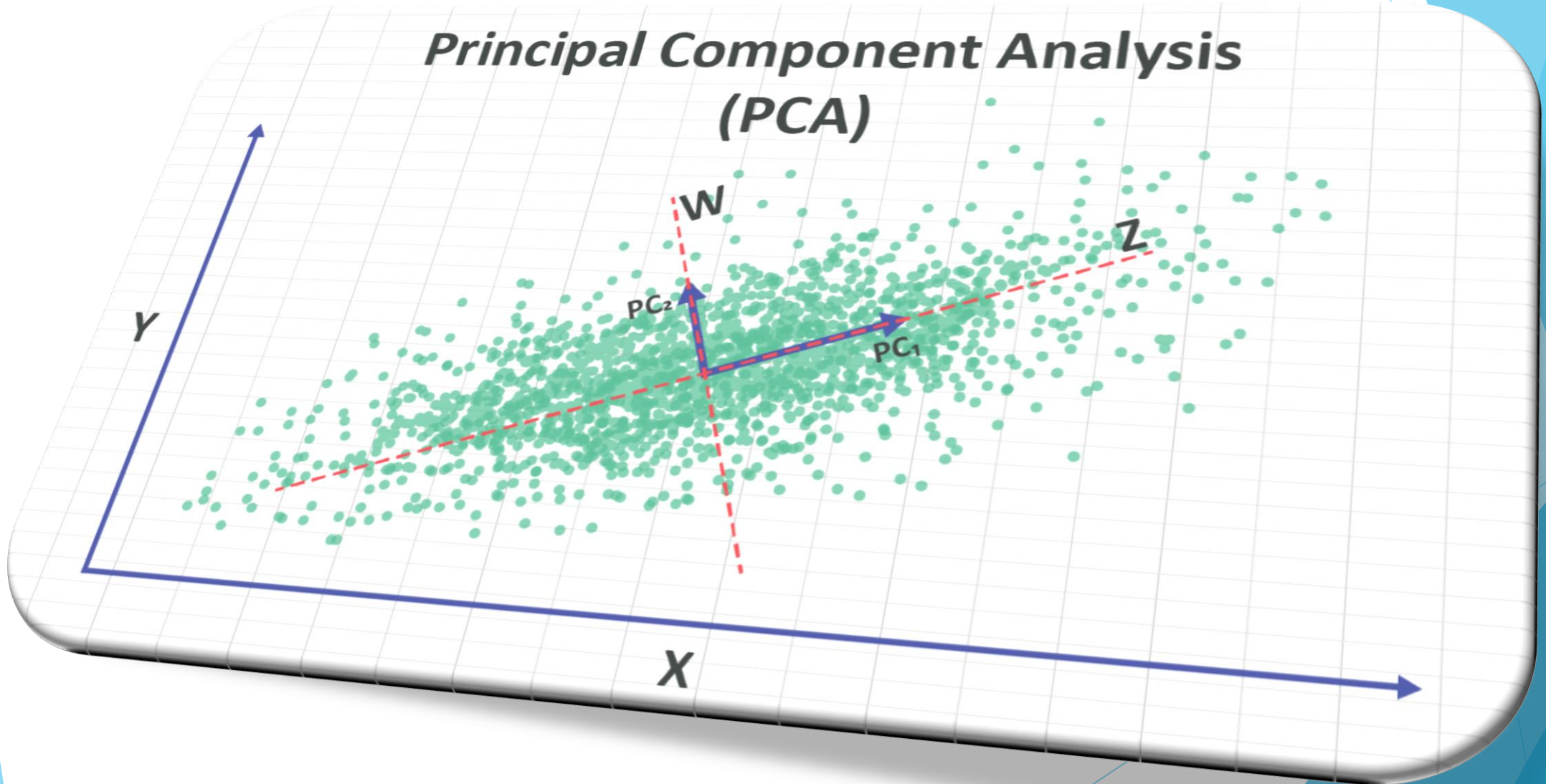


بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

المحور الرابع: تحليل المركبات الأساسية Principal Component Analysis



الهدف من المحور:



✓ لمحة تاريخية على تحليل المركبات الأساسية

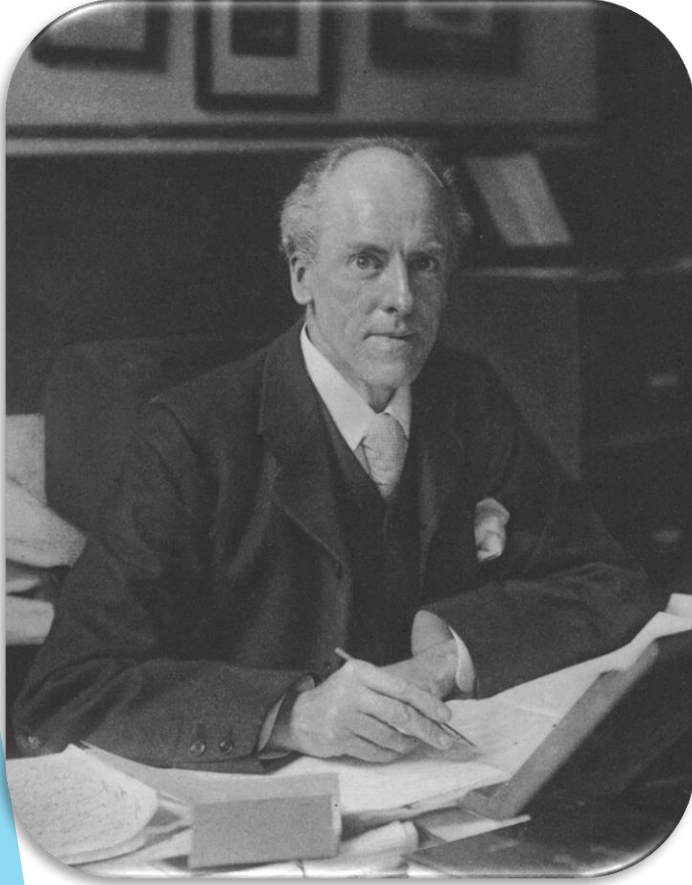
✓ تعريف تقنية تحليل المركبات الأساسية

✓ الهدف من تقنية تحليل المركبات الأساسية

✓ المبدأ الذي تقوم عليه تقنية تحليل المركبات الأساسية

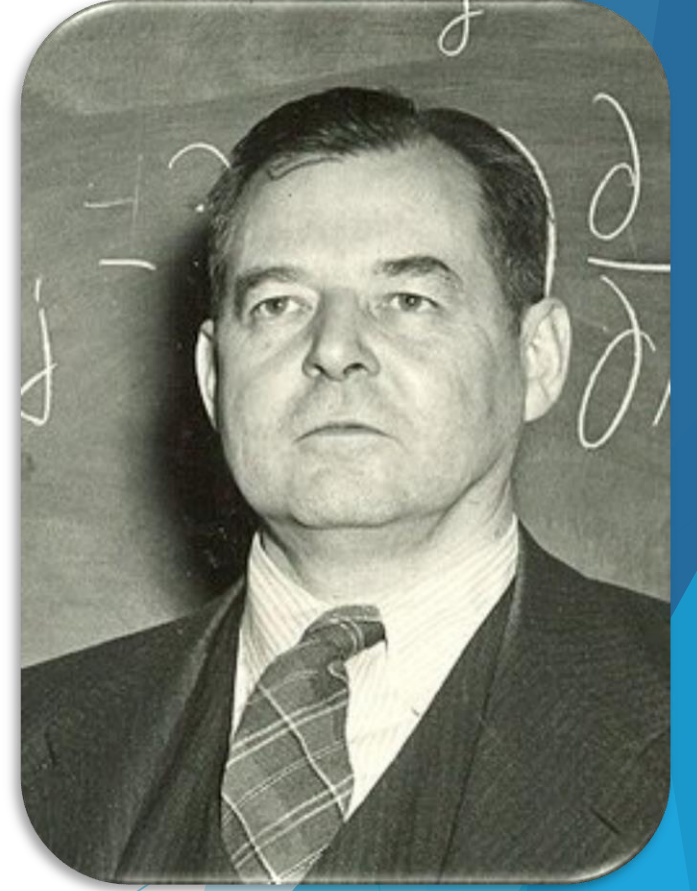
✓ خطوات تنفيذ تقنية تحليل المركبات الأساسية

1- لمحة تاريخية على تقنية تحليل المركبات الأساسية



استخدمت طريقة تحليل المركبات
الرئيسية لأول مرة من قبل Karl
pearson (1901)

ثم تم تطوير هذه الطريقة من قبل
Harold Hotelling وذلك عام
(1933)



2- تعريف تقنية تحليل المركبات الأساسية

كما يعرف تحليل المركبات الأساسية على أنها أحد تقنيات تحليل البيانات التي تختص في اختزال الجداول ذات الأبعاد الكبيرة، أي اختزال عدد كبير من المتغيرات الخام إلى عدد أقل من المتغيرات الجديدة والتي تعرف بالمركبات الأساسية.

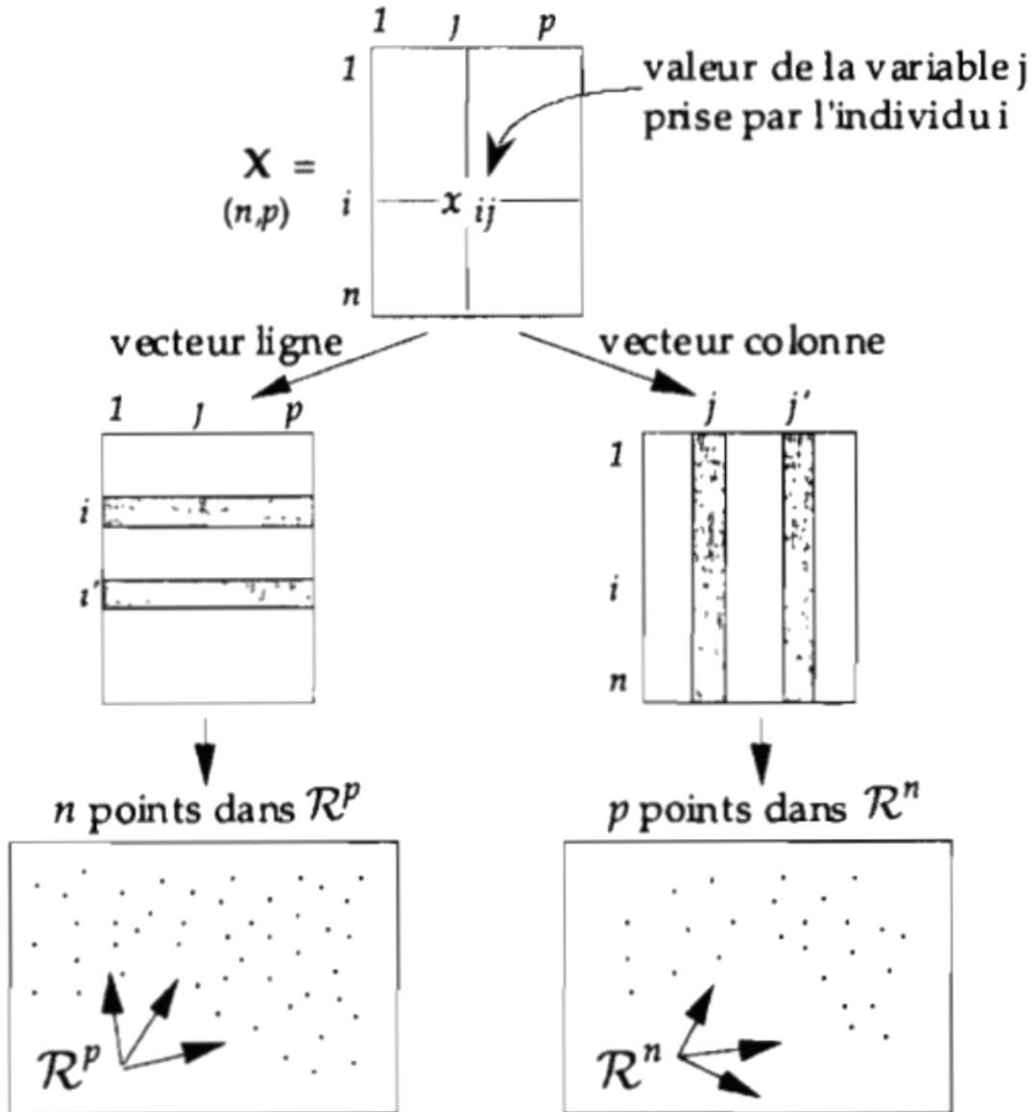
أسلوب رياضي يقوم على أساس تحويل مجموعة من المتغيرات المترابطة فيما بينها إلى مجموعة جديدة من المتغيرات غير المترابطة (أو المتعامدة - Orthogonal) تدعى بالمركبات الرئيسية

3- الهدف من استخدام تقنية تحليل المركبات الأساسية

تتضمن طريقة تحليل المركبات الأساسية هدفين رئيسيين:

- تلخيص (اختزال) جدول البيانات الذي يضم العديد من المتغيرات الكمية واختزالها في مركبتين أو ثلاث مركبات رئيسية
- التمثيل البياني لأكبر قدر ممكن من المعلومات الموجودة في جدول البيانات (والتي من الصعب تمثيلها في الأصل في فضاء متعدد الأبعاد)، في تمثيل بياني بسيط مكون من محورين أو ثلاث محاور.

4- مبدأ تقنية تحليل المركبات الأساسية



نقطة الانطلاق في التحليل بالمركبات الرئيسية، هو جدول X يحتوي على عدد p من المتغيرات تم قياسها على عدد n من الأفراد.

سحابة n من الأفراد تمثل في فضاء P من المتغيرات

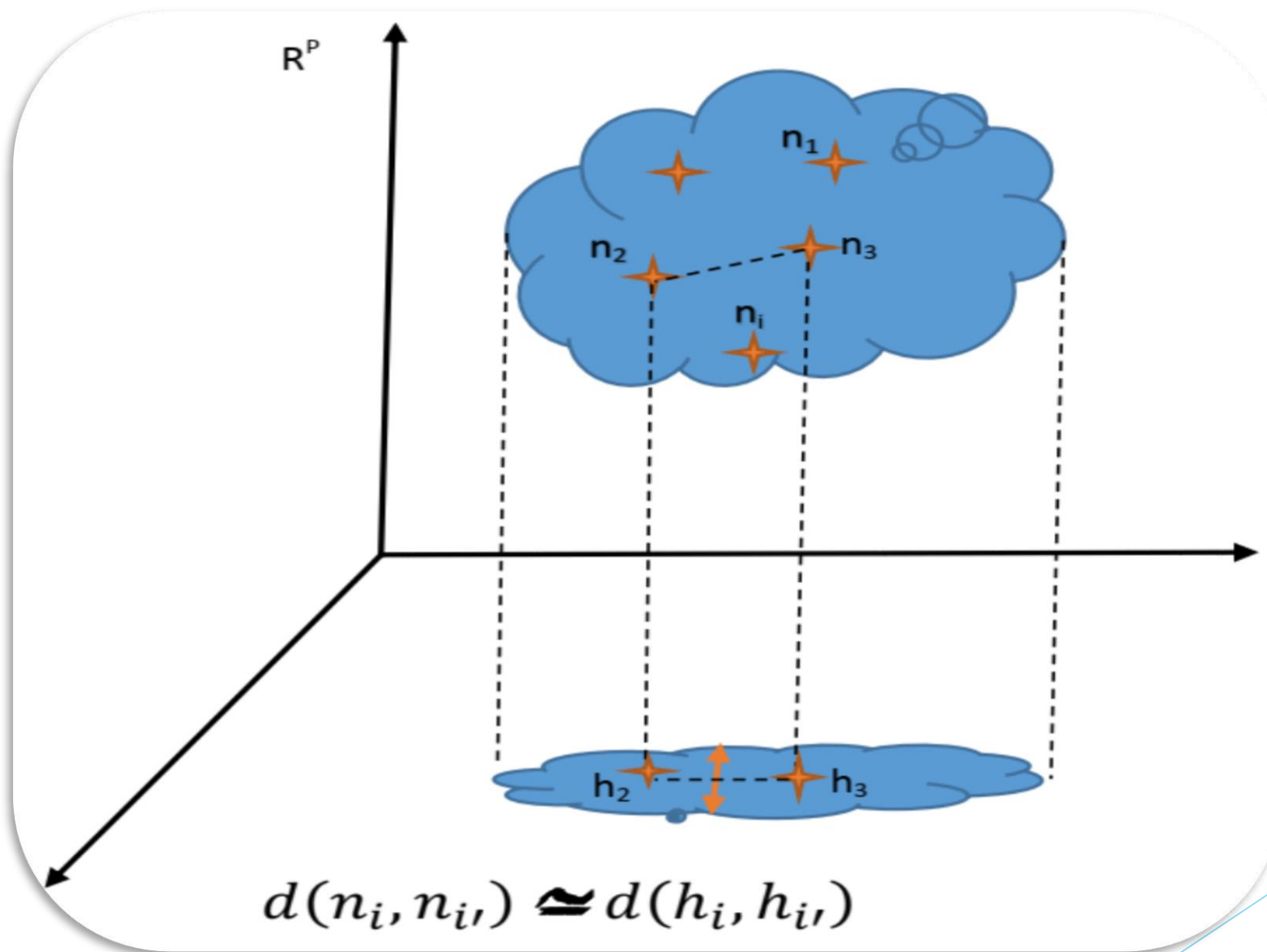
سحابة P من المتغيرات تمثل في فضاء n من الأفراد

إذا الجدول X هو مصفوفة حيث يمثل كل متجه (صف أو عمود) نقطة إما في فضاء R^p أو R^n

4- مبدأ تقنية تحليل المركبات الأساسية

- يتمثل المبدأ الأساسي لـ PCA (للتقليل من الأبعاد) في إسقاط سحابة نقاط الأفراد (المتغيرات) من فضاء ذو p (n) بُعد (متعدد الأبعاد) إلى فضاء جزئي ذو بعدين أو 3 أبعاد فقط. مع الحصول على تمثيل أقرب ما يمكن لسحابة النقاط الأصلية.
- أي العمل على عدم تشويه سحابة النقاط الأصلية عند إسقاطها
- بمعنى الحفاظ على المسافات بين نقاط السحابة الأصلية قدر الإمكان
- بتعبير آخر علينا أن نجعل المسافات الإجمالية بين النقاط المسقطة أقرب ما يمكن إلى المسافات الحقيقية بين النقاط في الفضاء الأصلي.

4- مبدأ تقنية تحليل المركبات الأساسية



4- مبدأ تقنية تحليل المركبات الأساسية



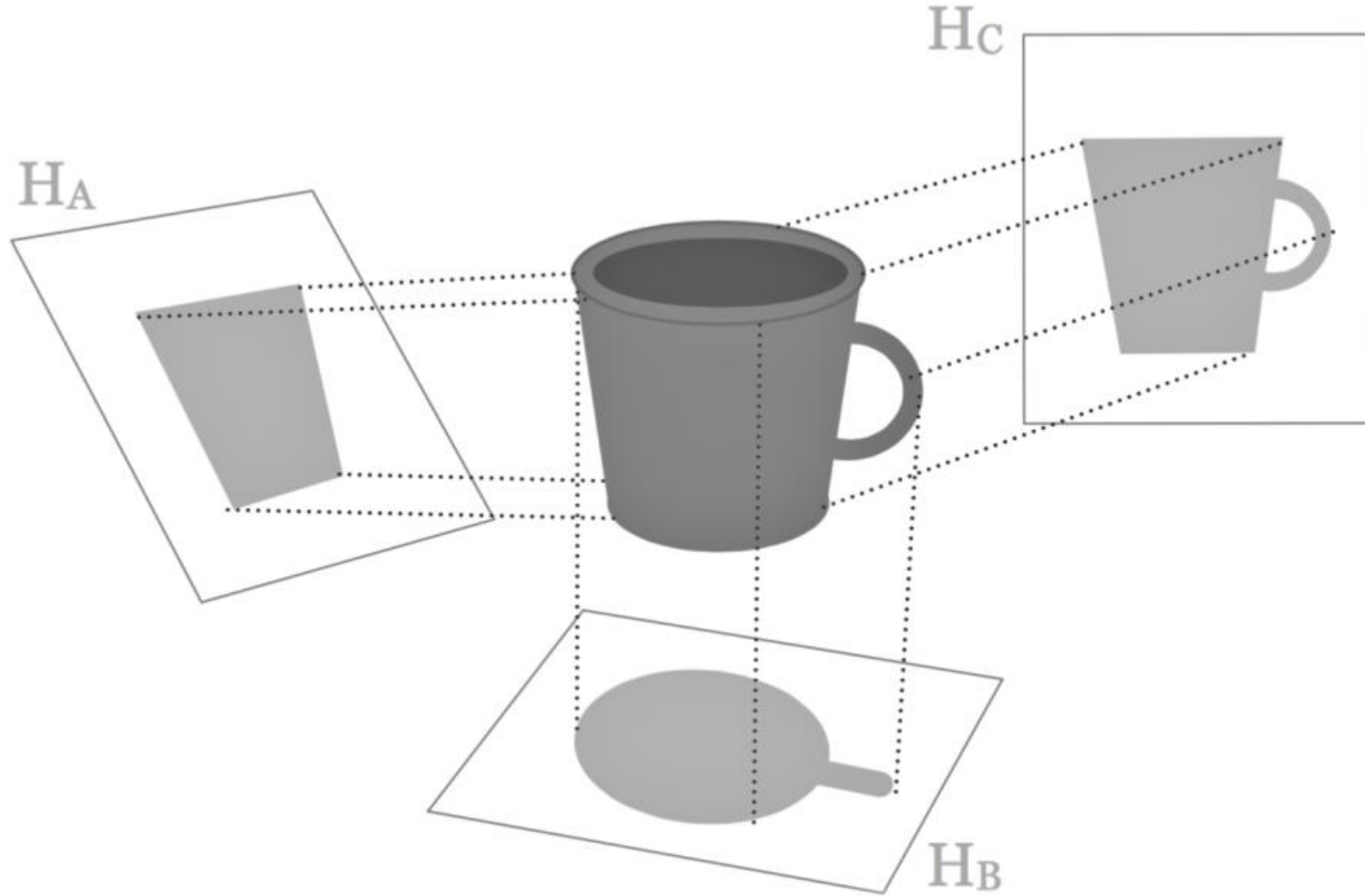
السؤال المطروح هو:

كيف نختار المحور (أو المستوى)؟ الذي يحقق لنا شرط

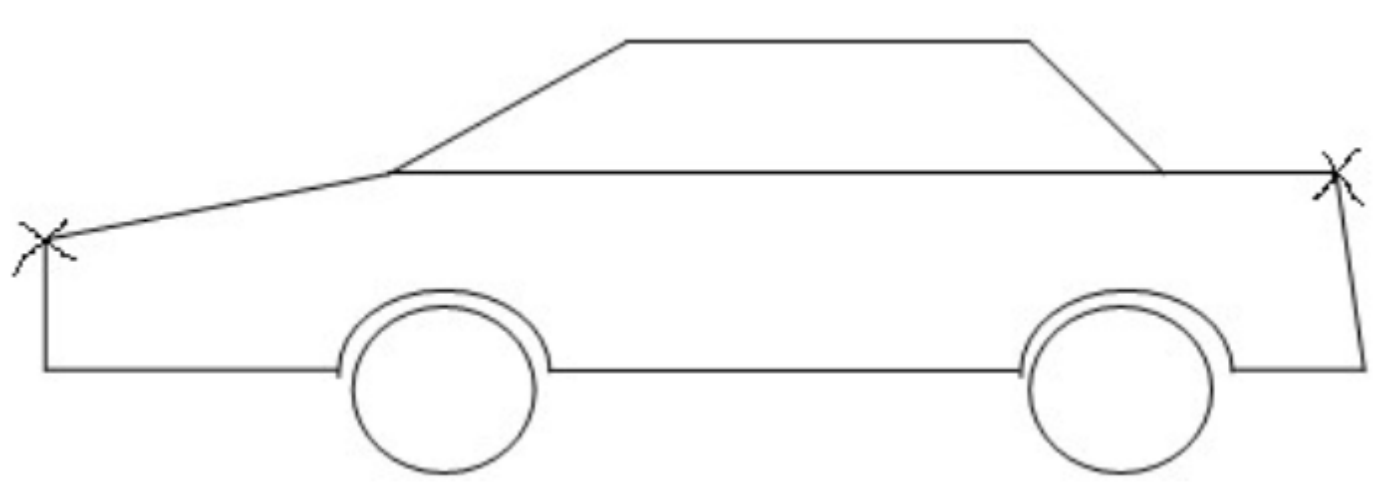
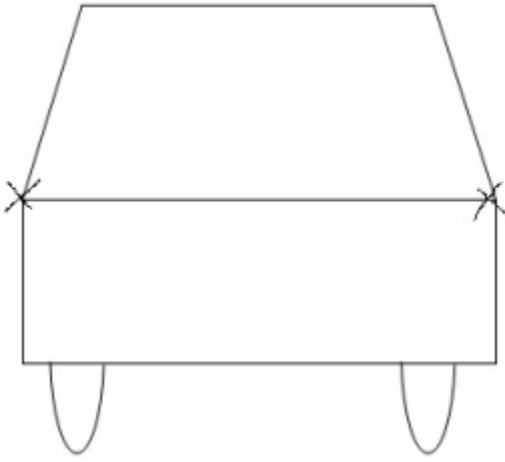
✓ المحافظة على المسافات بين النقاط قدر الإمكان

✓ أن تكون أكثر تمثيلا للسحابة الأصلية

4- مبدأ تقنية تحليل المركبات الأساسية



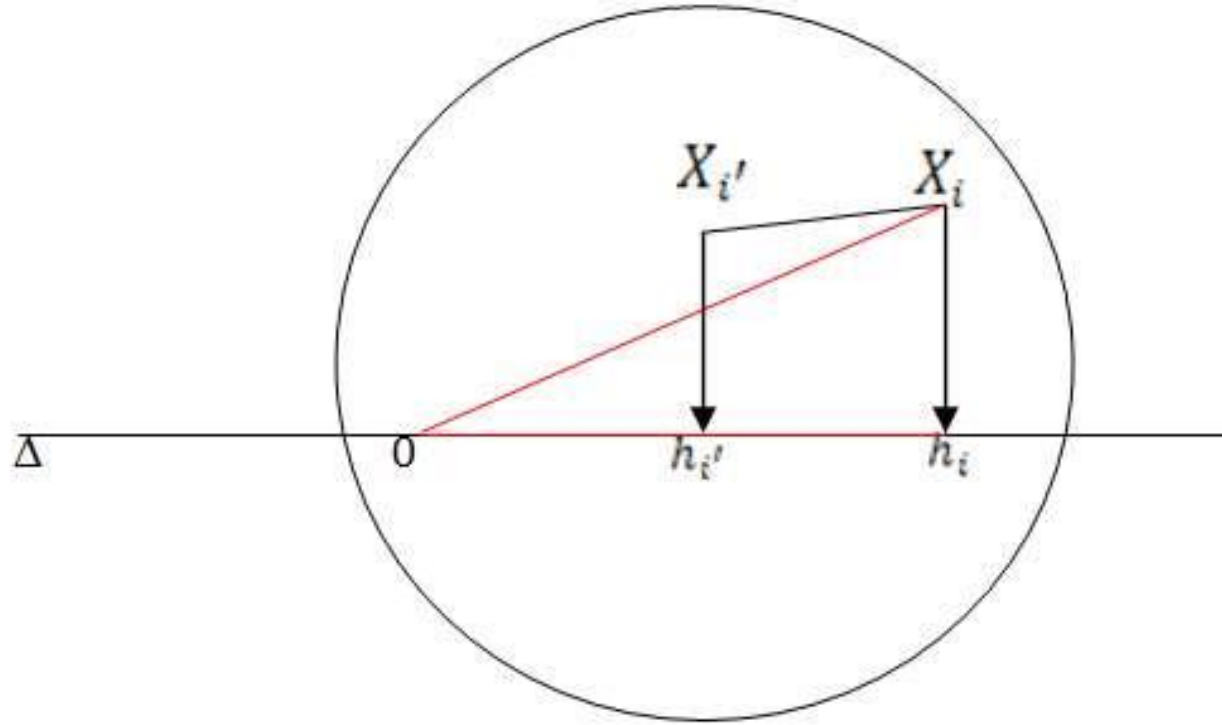
4- مبدأ تقنية تحليل المركبات الأساسية



✓ بالتالي يكون المحور (المستوي) الأكثر تمثيلا لسحابة النقاط هو الذي تكون فيه أطراف السحابة أكثر تشبها (أي أكبر تباين ممكن)

4- مبدأ تقنية تحليل المركبات الأساسية

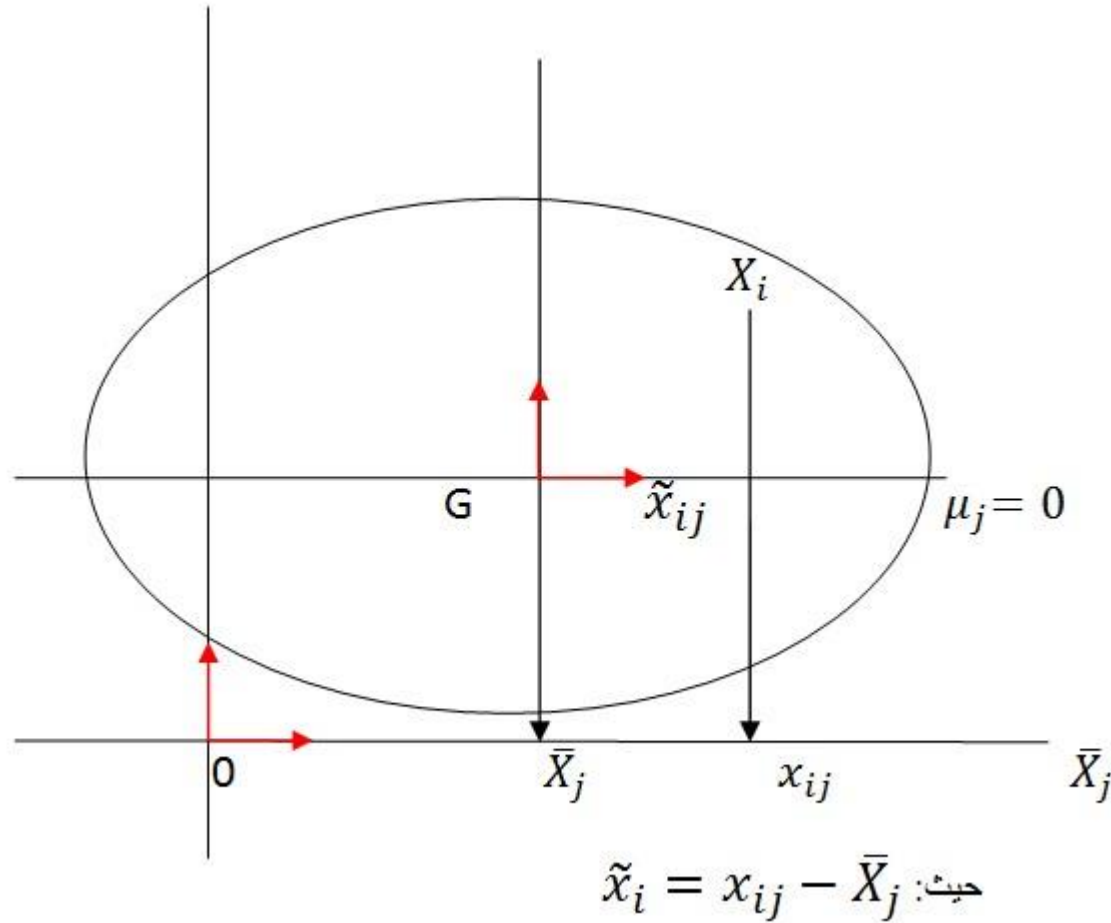
في التمثيل الهندسي يعتمد الإسقاط على مبدأ المحافظة على المسافات بين الأفراد



$$d(X_i, X_i') \approx d(h_i, h_i')$$

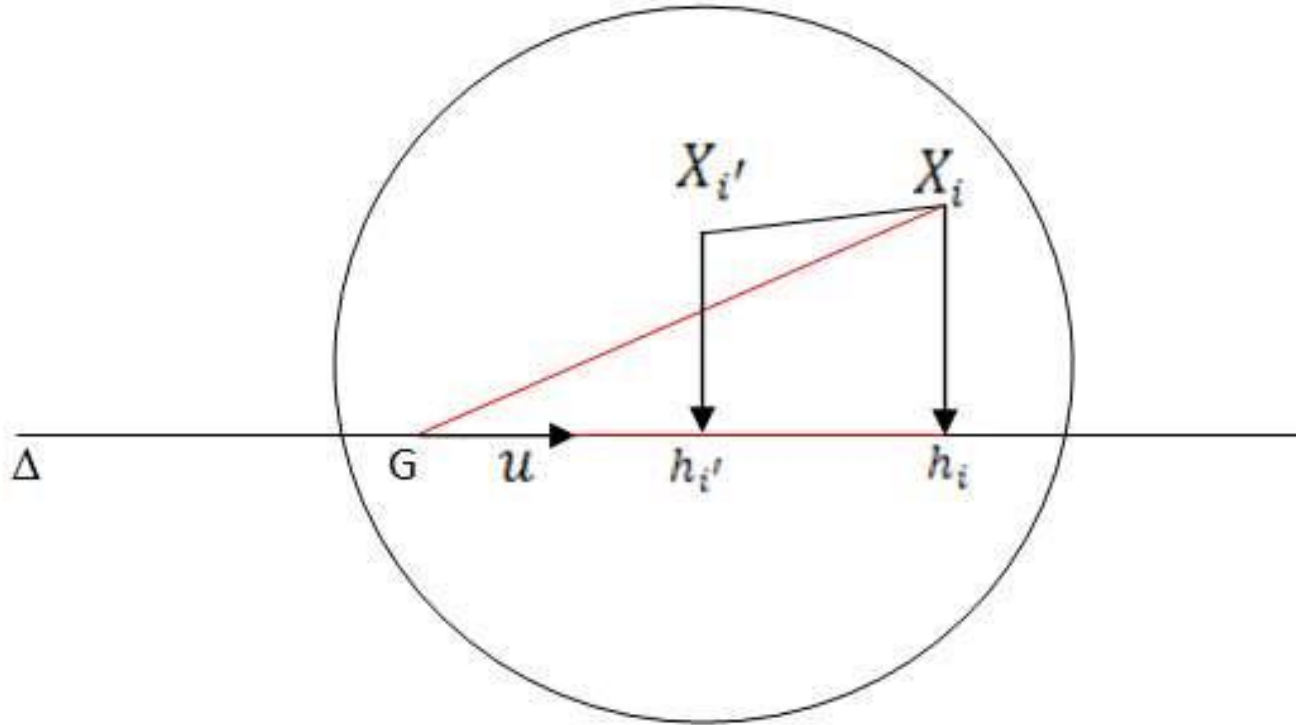
4- مبدأ تقنية تحليل المركبات الأساسية

عند الإسقاط على المستقيم Δ وللحصول على تمثيل أقل تشوه لسحابة النقاط، نأخذ مركز ثقل سحابة النقاط (G) كمركز للإسقاط، والذي يمثل متوسط المتغيرات



4- مبدأ تقنية تحليل المركبات الأساسية

وبشكل عام يمكن توضيح الإسقاط بجعل مركز الإسقاط هو النقطة G كما يلي:



$$Max(I_{\Delta}) = Max(\frac{1}{n} \sum_{i=1}^n d^2(G, h_i))$$

I_{Δ} (Inertie) هو تباين النقاط على طول المستقيم Δ (أي تشتت النقاط المسقطة على المستقيم Δ حول المركز G)

أن تعظيم التباين الكلي على طول المستقيم Δ يعني البحث عن أكبر القيم الذاتية لمصفوفة التباين والتباين المشترك/ أو لمصفوفة الارتباط (حسب طبيعة المتغيرات المدروسة)

5- أنواع تقنية تحليل المركبات الأساسية

يوجد نوعين (أو طريقتين):

الطريقة الأولى: تحليل المركبات الأساسية البسيطة (غير معياري) Non-Normalized

- ✓ تستخدم هذه الطريقة في حال كانت جميع المتغيرات متجانسة
- ✓ يتم العمل في هذه الحالة على جدول البيانات الممركز
- ✓ وبالتالي نستخدم مصفوفة التباين والتباين المشترك

5- أنواع تقنية تحليل المركبات الأساسية

الطريقة الثانية: تحليل المركبات الأساسية المعيارية Normalized

- ✓ تستخدم هذه الطريقة في حال كانت المتغيرات غير متجانسة
- ✓ يتم العمل في هذه الحالة على جدول البيانات المعيارية (المركزة المختزلة)
- ✓ وبالتالي نستخدم مصفوفة الارتباط

5- خطوات إجراء تقنية تحليل المركبات الأساسية

1-6 تشكيل جدول البيانات الكمية:

- ✓ يستخدم أسلوب تحليل المركبات الرئيسية مع جدول البيانات الكمية
- ✓ يضم أفراد وكذلك متغيرات كمية مستمرة
- ✓ قد تكون المتغيرات متجانسة أو غير متجانسة في معيار القياس
- ✓ تكون المتغيرات مترابطة فيما بينها

6- خطوات إجراء تقنية تحليل المركبات الأساسية

مثال: ليكن لدينا نقاط خمس (5) طلبة في ثلاث (3) مقاييس.

المتغيرات	X ₁	X ₂	X ₃
الأفراد			
1	11	10	13
2	11	8	11
3	14	13	15
4	10	8	11
5	9	11	10

$$X_{5 \times 3} = \begin{pmatrix} 11 & 10 & 13 \\ 11 & 8 & 11 \\ 14 & 13 & 15 \\ 10 & 8 & 11 \\ 9 & 11 & 10 \end{pmatrix}$$

6- خطوات إجراء تقنية تحليل المركبات الأساسية

2-6 حساب المتوسطات وتشكيل جدول البيانات الكمية الممركز:

- ✓ حساب المتوسطات الحسابية لكل متغير على حدة (كل عمود لديه متوسط حسابي)
- ✓ طرح المتوسطات الحسابية من القيم x_i الموجودة في نفس العمود
- ✓ و G هي نقطة مركز ثقل سحابة النقاط إحداثياتها هي متوسطات المتغيرات
- ✓ جعل البيانات ممركرة لا يؤثر على شكل سحابة النقاط

المتغيرات الأفراد	X ₁	X ₂	X ₂
1	11	10	13
2	11	8	11
3	14	13	15
4	10	8	11
5	9	11	10
G	11	10	12

$$\bar{X}_j = \frac{\sum_{i=1}^n x_i}{n}$$

يمكن حسابها لكل عمود بالقانون:

$$\bar{X}_1 = \frac{11+11+14+10+9}{5} = 11$$

لدينا: $\bar{X}_1 = 11$ ، $\bar{X}_2 = 10$ ، $\bar{X}_3 = 12$.

جدول البيانات الممركزة

المتغيرات الأفراد	X ₁	X ₂	X ₂
1	11-11=0	10-10=0	13-12=1
2	0	-2	-1
3	3	3	3
4	-1	-2	-1
5	-2	1	-2

المصفوفة الممركز

$$\Rightarrow X_c = \begin{pmatrix} 0 & 0 & 1 \\ 0 & -2 & -1 \\ 3 & 3 & 3 \\ -1 & -2 & -1 \\ -2 & 1 & -2 \end{pmatrix}$$

يمكن حسابها
بطرح المتوسط
الحسابي من كل
عنصر في العمود
الذي يوافقه

6- خطوات إجراء تقنية تحليل المركبات الأساسية

3-6 حساب التباينات والانحرافات وتشكيل جدول البيانات المعيارية (المركزة المختزلة)

✓ بالنسبة للبيانات الغير متجانسة في وحدات القياس يجب تحويلها الى بيانات معيارية

✓ نحسب أولا التباينات ثم الانحرافات المعيارية (لكل عمود) $v = \frac{\sum (x_i - \bar{X})^2}{n}, s = \sqrt{v}$

✓ نقسم عناصر كل عمود من المصفوفة المركزة على الانحراف المعياري المقابل له

جدول البيانات الممركزة

المتغيرات الأفراد	X ₁	X ₂	X ₂
1	0	0	1
2	0	-2	-1
3	3	3	3
4	-1	-2	-1
5	-2	1	-2
V	2.8	3.6	3.2
S	1.67	1.89	1.78



جدول البيانات المعيارية

المتغيرات الأفراد	X ₁	X ₂	X ₂
1	0	0	0.56
2	0	-1.05	-0.56
3	1.79	1.58	1.68
4	-0.6	-1.05	-0.56
5	-1.2	0.53	-1.12



المصفوفة المعيارية

$$X_{CR} = \begin{pmatrix} 0 & 0 & 0,56 \\ 0 & -1,05 & -0,56 \\ 1,79 & 1,58 & 1,68 \\ -0,6 & -1,05 & -0,56 \\ -1,2 & 0,53 & -1,12 \end{pmatrix}$$

6- خطوات إجراء تقنية تحليل المركبات الأساسية

4-6 حساب مصفوفة التباين والتباين المشترك أو حساب مصفوفة الارتباط

1-4-6 حساب مصفوفة التباين والتباين المشترك

✓ نحسب مصفوفة التباين والتباين المشترك V في حال كانت المتغيرات متجانسة

✓ يتم حساب التباين والتباين المشترك من الصيغة المصفوفية التالية:

$$V = \frac{1}{n} \cdot X_C^T \cdot X_C \quad \text{أو} \quad \text{حيث:}$$

X_C^T هي منقول المصفوفة الممركزة

X_C هي المصفوفة الممركزة

6- خطوات إجراء تقنية تحليل المركبات الأساسية

ملاحظات:

✓ مصفوفة التباين والتباين المشترك V هي مصفوفة متناظرة ومن الدرجة $p \times p$

✓ أن القطر الرئيسي للمصفوفة V هي تباينات المتغيرات

✓ المصفوفة V ، مصفوفة متماثلة (تساوي منقولها) $V = V^T$

✓ عند استخدام المصفوفة V ، يحسب التباين الكلي (Inertie Total) من العلاقة:

$$I_T = \text{Trace}(V) = \sum \text{Var}$$

حساب مصفوفة التباين والتباين المشترك

$$V = \frac{1}{n} \cdot X_C^T \cdot X_C = \frac{1}{5} \cdot \begin{pmatrix} 0 & 0 & 3 & -1 & -2 \\ 0 & -2 & 3 & -2 & 1 \\ 1 & -1 & 3 & -1 & -2 \end{pmatrix} \cdot \begin{pmatrix} 0 & 0 & 1 \\ 0 & -2 & -1 \\ 3 & 3 & 3 \\ -1 & -2 & -1 \\ -2 & 1 & -2 \end{pmatrix} = \begin{pmatrix} 2.8 & 1.8 & 2.8 \\ 1.8 & 3.6 & 2.2 \\ 2.8 & 2.2 & 3.2 \end{pmatrix}$$

6- خطوات إجراء تقنية تحليل المركبات الأساسية

4-6 حساب مصفوفة التباين والتباين المشترك أو حساب مصفوفة الارتباط

2-4-6 حساب مصفوفة الارتباط:

✓ نحسب مصفوفة الارتباط R في حال كانت المتغيرات متجانسة

✓ يتم حساب مصفوفة الارتباط من الصيغة المصفوفية التالية:

$$R = \frac{1}{n} \cdot X_{CR}^T \cdot X_{CR} \quad \text{أو}$$

حيث:

X_{CR}^T هي منقول المصفوفة المعيارية

X_{CR} هي المصفوفة المعيارية

6- خطوات إجراء تقنية تحليل المركبات الأساسية

ملاحظات:

✓ مصفوفة الارتباط R هي مصفوفة متناظرة ومن الدرجة $p \times p$

✓ أن كل أعداد القطر الرئيسي للمصفوفة R هي الواحد الصحيح (1)

✓ المصفوفة R ، مصفوفة متماثلة (تساوي منقولها) $R = R^T$

✓ عند استخدام المصفوفة R ، يحسب التباين الكلي (Inertie Total) من العلاقة:

$$I_T = \text{Trace}(R) = p$$

حساب مصفوفة الارتباط:

$$R = \frac{1}{n} \cdot X_{CR}^T \cdot X_{CR} = \frac{1}{5} \cdot \begin{pmatrix} 0 & 0 & 1.79 & -0.6 & -1.2 \\ 0 & -1.05 & 1.58 & -1.05 & 0.53 \\ 0.56 & -0.56 & 1.68 & -0.56 & -1.12 \end{pmatrix} \cdot \begin{pmatrix} 0 & 0 & 0.56 \\ 0 & -1.05 & -0.56 \\ 1.79 & 1.58 & 1.68 \\ -0.6 & -1.05 & -0.56 \\ -1.2 & 0.53 & -1.12 \end{pmatrix} = \begin{pmatrix} 1 & 0.57 & 0.94 \\ 0.57 & 1 & 0.65 \\ 0.94 & 0.65 & 1 \end{pmatrix}$$