جامعة محمد خيضر (بسكرة)
كلية العلوم الاقتصادية والتجارية والتسيير
قسم العلوم التجارية

السنة الجامعية: 2025/2024
المستوى: أولى ماستر تسويق
مقياس: الأساليب الكمية في التسويق 2

# المحاضرة الخامسة: تحليل المركبات الرئيسية

# Principal Component Analysis (PCA)- Analyse en composantes principales (ACP)

## 1. تمهيد

تعتبر طريقة تحليل المركبات أو المكونات الرئيسية (PCA) من أشهر الطرق واكثرها استعمالا في مختلف البحوث والدراسات العلمية في مجالات عدة (علم الحاسوب، الطب، البيولوجيا، .... )، وهذا بسبب قدرتها الكبيرة على ضغط وتقليص البيانات، مما يجعلها مفيدة جدا عندما يكون هناك كم هائل من البيانات الكمية القابلة للمعالجة لغرض التفسير. وتستخدم طريقة (PCA) بشكل شائع في التسويق لتجميع عناصر من أداة القياس (فقرات الاستبيان مثلا) في أبعاد أو عوامل قليلة متماسكة.

## 2. نشأة تحليل المركبات الرئيسية

تم اختراع PCA عام 1901 من قبل كارل بيرسون Karl Pearson، منظر اختبار كاي مربع ($\chi^2$)، محاولةً منه لتعميم مبادئ نظرية عزم القصور الذاتي في الميكانيكا، وقد حاول بيرسون، آنذاك، تعميم أعماله في مجالي الانحدار والارتباط، ليس لتفسير متغير بآخر، بل لوصف البيانات وتلخيصها.

وفي 1930، قام هارولد هوتلين Harold Hotelling بتسمية وتطوير الطريقة، في إطار تقعيد نظري شمل طرق التحليل العاملي التي انبثقت من طريقة تحليل المركبات الرئيسية، وقد كان أكبر عائق أمام تبني هذه الطريقة في النصف الأول من القرن العشرين، هو الكم الهائل من الحسابات المعقدة التي تستهلكها، والتي لم تكن إمكانيات الحساب الآلي المتوفرة آنذاك تسمح بها.

وابتداء من الستينات، ومع تطور الحساب الآلي، تنامت شعبية طريقة المركبات الأساسية وباقي طرق التحليل العاملي في مجتمع الإحصائيين، بل وتم تبنيها من طرف مجالات معرفية متنوعة أخرى كالاقتصاد، علم الاجتماع والمعلوماتية بهدف المعالجة الرقمية وضغط البيانات.

## 3. تعريف تحليل المركبات الرئيسية

تحليل المركبات الرئيسية هو: "أسلوب رياضي يقوم على أساس تحويل مجموعة من المتغيرات التفسيرية المترابطة فيما بينها إلى مجموعة جديدة من المتغيرات المستقلة (المتعامدة Orthogonal) تدعى المركبات الرئيسية، حيث كل مركبة رئيسية هي عبارة عن توليفة خطية للمتغيرات الأصلية" .

تحليل المكونات الرئيسية هو: "تقنية تقليل الأبعاد التي تحول مجموعة كبيرة من المتغيرات إلى مجموعة أصغر، مع الاحتفاظ بأهم المعلومات في المجموعة الأصلية، أي أن الهدف الرئيسي من طريقة PCA هو تقليل تعقيد البيانات، دون فقدان جوهر ما تمثله هذه البيانات".

## 4. بعض استخدامات تحليل المركبات الرئيسية

- **في التمويل**، تُستخدم PCA لتقليل أبعاد مجموعة كبيرة من الأصول المالية (مثل الأسهم) إلى عدد أقل من المكونات التي تلتقط اتجاهات السوق الأساسية، ويمكن للمستثمرين استخدام هذه المكونات لتقييم المخاطر وتخصيص الأصول بشكل أكثر فعالية.
- **في التصوير الطبي**، يمكن استخدام PCA لتقليل تعقيد الصور الطبية مع الاحتفاظ بالمعلومات الهامة، فيساعد ذلك في تحسين جودة التصوير وتشخيص الأمراض بشكل أكثر دقة.
- **في التصنيع**، يمكن للشركات المصنعة استخدام PCA لمراقبة الجودة، من خلال تحليل عدد كبير من العوامل التي تؤثر على الإنتاج، وعبر تقليل هذه العوامل، يمكن تحديد المشكلات المحتملة بشكل أسرع واتخاذ الإجراءات التصحيحية.
- **في التسويق**، يمكن للشركات استخدام PCA لتحليل أنماط سلوك العملاء، من خلال تقليل عدد المتغيرات مثل الدخل، العمر، التفضيلات...، إلى عدد أقل من المكونات التي تفسر بشكل أفضل سلوكيات الشراء.

**5. خصائص طريقة تحليل المركبات الرئيسية**

**أ. طريقة وصفية (Descriptive):** تقوم بتحويل المعطيات الأولية إلى مخططات ورسومات، وبالتالي تمنحها الطابع الوصفي والمرئي، عبر تقليص الأبعاد (المغيرات) إلى بعدين مثلا، وبذلك يتم تمثيل البيانات على ورقة في المستوي، على أن يحتفظ الرسم البيانأكبر قدر ممكن من المعلومات الموجودة في جدول البيانات.

**ب. طريقة استكشافية (Exploratoire):** تهدف طريقة PCA استكشاف الظواهر عبر إظهار الفجوات بين المتغيرات وتشكيل مجموعات متشابهة من المفردات، وليس اختبار فرضيات أو نماذج مسبقة (طريقة استدلالية deductive). ففي حالة الاستبيان، لا تقوم PCA باختبار علاقة الرضا بالولاء مثلا، وإنما تقوم بدراسة مجموع فقرات الاستبيان (المتغيرات)، بهدف اختزالها وتجميعها في أبعاد أو محاور مستقلة .

**ج. طريقة متعددة الأبعاد (Multidimensionnelle):** تنطلق طريقة PCA دائما من عدة متغيرات كمية يتم قياسها على عينة من المفردات المعنيين، ثم اختزالها في مركبتين (بعدين) أو أكثر، ففي حالة الاستبيان مثلا فإن طريقة PCA لا تدرس إجابة واحدة، وإنما جميع الإجابات المشكلة له باعتبار كل سؤال هو متغير في حد ذاته .

**د. طريقة تلخيصية- اختزالية:** تهدف طريقة PCA إلى تلخيص جدول البيانات الذي يضم العديد من المتغيرات الكمية، واختزالها في مركبتين أو ثلاث مركبات رئيسية (تسمى عوامل أو مكونات)، وهذا لا يعني أبدا حذف المتغيرات، مما يجعلها تسمح بتعيين الأفراد المتشابهين والمختلفين، وتحديد درجة الترابط بين المتغيرات، واكتشاف المتغيرات المسؤولة عن التشابه أو الاختلاف بين الأفراد.

**هـ. طريقة رياضية:** يتمثل مبدأ ACP في إسقاط سحابة نقاط الأفراد (فضاء $R^n$ ذو بعد n)، والمتغيرات (فضاء $R^p$ ذو بعد p)، إلى محاور أو مستويات من 2 إلى 3 أبعاد فقط (فضاء شعاعي جزئي ذو بعد أقل)، مع الحفاظ قد المستطاع على المسافات بين الأفراد والمتغيرات قدر الإمكان، أي بتعبير آخر التقليل أقصى ما يمكن من التشوه الذي قد يطرأ على الساحبة الأصلية نتيجة الإسقاط (الحصول على تباين كلي أكبر ما يمكن).

**و. طريقة تعظيمية للتباين:** على الرغم من أنه يتم استخراج المركبة الأولى لالتقاط أقصى قدر من التباين الكلي، إلا أنه نادرا ما يمكن التقاط التباين الكلي. وبالتالي، فإن ما تبقى من تباين تولده المركبة الثانية ... وهكذا، ومع ذلك لا يمكن أن يتجاوز عدد المركبات المستخرجة عدد المتغيرات الأصلية، كما يمكن الاقتصار على عدد محود من المركبات (2 أو 3) التي تحتافظ بأكبر نسبة من التباين الكلي.

**6. أمثلة واقعية**

المكونات الرئيسية هي متغيرات جديدة يتم إنشاؤها كتركيبات خطية للمتغيرات الأصلية في مجموعة البيانات. هذه المكونات غير مترابطة وتُرتب حسب مقدار التباين الذي تلتقطه من البيانات الأصلية. المكون الرئيسي الأول يلتقط أكبر قدر من التباين، يليه المكون الثاني، وهكذا. ولفه ذلك نستعرض بعض الأمثلة من الواقع.

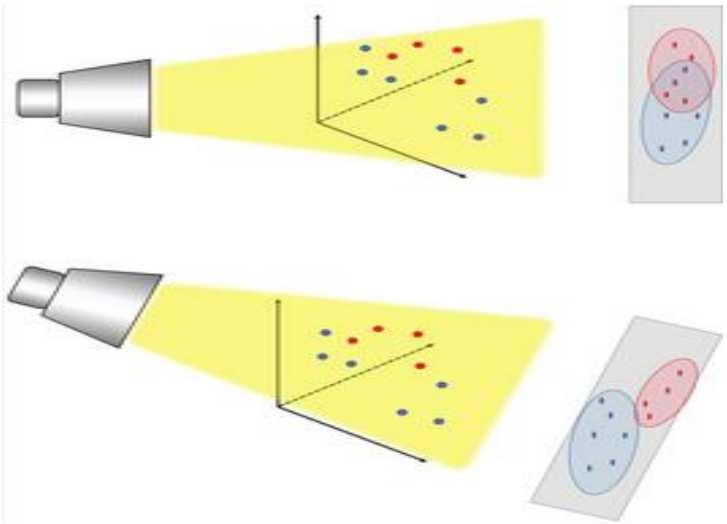
**مثال (1):**

لدينا مجموعة بيانات تحتوي على 10 خصائص (متغيرات)، يمكن لتحليل المركبات الرئيسية تقليل هذا العدد إلى 3 خصائص (متغيرات) مستقلة فقط على سبيل المثال، شرط أن تبقى تحتفظ بمعظم التباين والأنماط الموجودة في مجموعة البيانات الأصلية، وهذا التبسيط يجعل استكشاف البيانات وتحليلها وتمثيلها بيانيا أسهل.

**مثال (2):**

مجموعة بيانات تحتوي على ثلاثة متغيرات: الطول، الوزن، والعمر. هذه المتغيرات مترابطة (بشكل عام كلما زاد طول، فإن الوزن يزداد)، مما يعني أن التغييرات في أحدها قد تؤثر على باقي المتغيرات. هنا يقوم تحليل المركبات الرئيسية PCA بإنشاء متغيرات جديدة أقل وغير مترابطة (تسمى المكونات الرئيسية) تلتقط جوهر ومعظم البيانات الأصلية. على سبيل المثال: المركبة الرئيسية الأولى $PC_1$ قد يلتقط الحجم الإجمالي للشخص، من خلال دمج الطول والوزن. المركبة الرئيسية الثانية $PC_2$ قد يلتقط التباين المرتبط بالعمر لوحده. ومن خلال التركيز على هذه المكونات الرئيسية، يمكننا تقليل أبعاد مجموعة البيانات مع الاحتفاظ بمعظم المعلومات (بدل 3 متغيرات نستخدم بعدين فقط).

**مثال (3):**



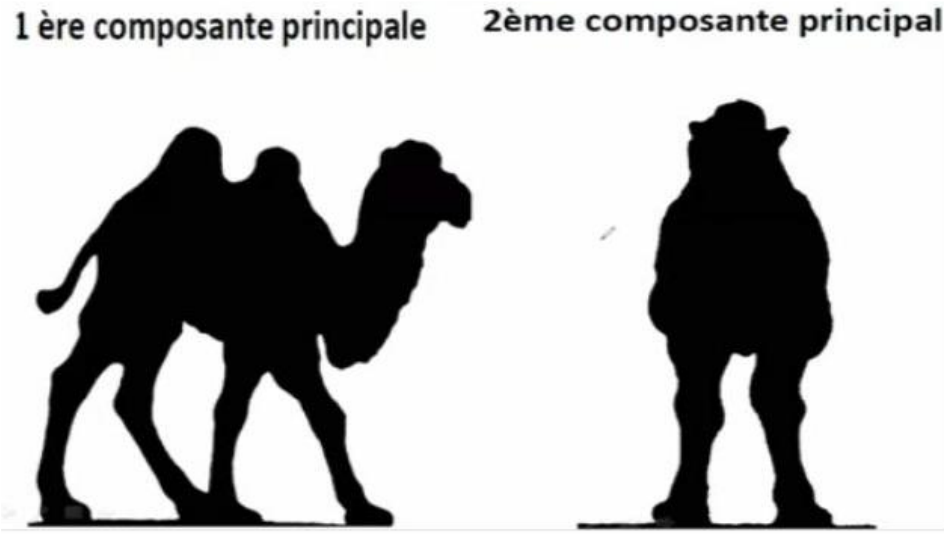
لنفترض تسليط ضوء على مجموعة من النقاط من مجسم ثلاثي الأبعاد، وملاحظة ظلال هذه النقاط في بُعدين على الحائط. يُمثل تسليط الضوء على البيانات البحث عن المستوى الذي يُحدث أكبر تباين بين مجموعات البيانات. تُشير زاوية الضوء في الصورة السفلية إلى مستوى ثنائي الأبعاد أفضل لإسقاط النقاط عليه مقارنةً بالصورة أدناه.

**مثال (3):**

يهتم أحد الباحثين بدراسة سلوكيات طلاب الماستر، فيجمع بيانات عينة كبيرة من الطلاب حول الدافعية الذاتية، القدرة الفكرية، التاريخ الدراسي، التاريخ العائلي، الحالة الصحية، الخصائص الجسمانية، ... الخ، ويقوم بقياس كل من هذه المجالات بعدة متغيرات، ثم يقوم بإدخال المتغيرات في التحليل ويدرس الارتباطات فيما بينها.

يكشف التحليل عن أنماط الارتباط بين المتغيرات التي يعتقد أنها تعبر عن الأبعاد الرئيسية التي تؤثر على سلوكيات الطلاب، فمثلا تتحد العديد من متغيرات القدرة الفكرية مع بعض متغيرات التاريخ الدارسي لتشكيل عامل الذكاء، وبالمثل قد تتحد متغيرات الدافعية والتاريخ الدراسي لتشكيل عامل تفضيل الاستقلالية

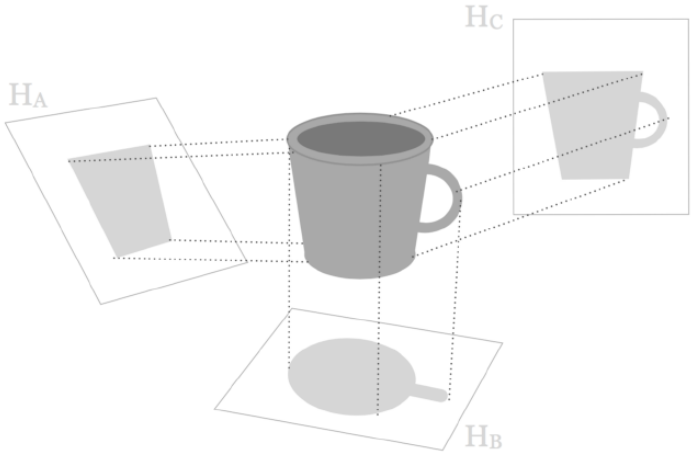
**مثال (4)**

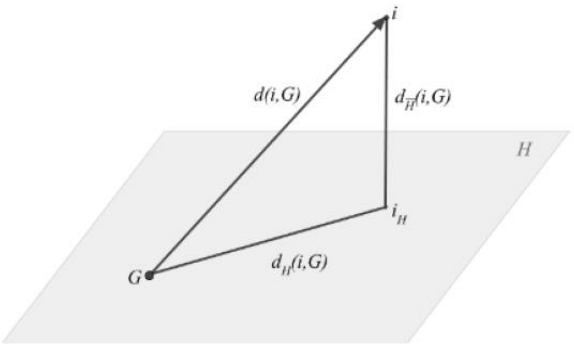



لو نظرنا فقط للصورة الموجودة على اليمين (2$^{éme}$ composante principal) فهل سنكون متأكدين بنسبة مائة في المائة أنها صورة جمل، بالطبع لا، ربما تكون صورة لحصان أو إنسان متخفي...الخ. سنكون متأكدين بنسبة 20 % أنّه جمل. الآن لو ننظر إلى الصورة على اليسار(1$^{ère}$ composante principal)، يمكننا التأكد بنسبة 70 % من أنه جمل. واذا نظرنا للصورتين معا (استخدام المركبين معا)، سنكون متأكدين بنسبة 90 % أن الصورة لجمل.

إنّ النتيجة المهمة من هذه التجربة أنه عند انتقالنا من رؤية ثلاثية الأبعاد إلى رؤية ثنائية الأبعاد سنحتفظ بتسعين في المائة من المعلومات، والعشرة في المائة الضائعة لم تأثر على قرارنا في أن الصورة لجمل. بمعنى آخر أنّ تقليل عدد الأبعاد مع خسارة طفيفة في المعلومات يكون أفضل لنا لاتخاذ القرارات الصائبة.

**مثال (5):**

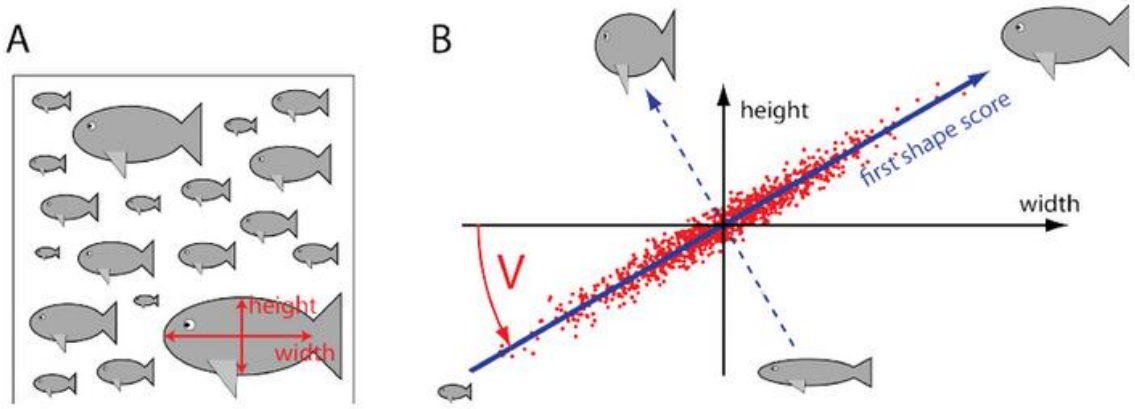



بالنظر إلى الصورة على اليسار، يمكن القول أنّ الإسقاط على المستوى $H_A$ أكثر فائدة من الإسقاط على المستوى $H_B$، فعلى الأقل يمكننا أن نرى أن الشكل في $H_A$ له علاقة بجسم طويل، وأن أحد نهاياته أوسع من الأخرى، وعلى النقيض من ذلك، فإن جميع نقاط السحابة المسقطة على المستوى $H_B$ مشوشة، ولا تنقل فكرة واضحة عن السحابة الأصلية باستثناء ظل المقب، وأفضل إسقاط بين الاسقاطات الثلاث هو الاسقاط $H_C$.

ولكي نحصل على الاسقاط $H_C$، يجب البحث عن مستوى الإسقاط الذي يجعل تباين نقاط السحابة أكبر ما يمكن، ونكتب: $\mathbf{Max_H \Sigma_i \Sigma_{i'} d_H{}^2 (i , i')}$

**مثال (6)**




**الرسم على اليسار (A):** يعطي مجموعة بيانات افتراضية لقياسات العرض والارتفاع لمجموعة من n مفردة (أسماك)، أي أن هناك فقط m = 2 من السمات الهندسية المقاسة هنا، نلاحظ من أحجام الأسماك أن هناك علاقة ارتباط بين عرض وارتفاع بين المفردات (الأسماك).

**الرسم على اليمين (B):** بما العرض والارتفاع مرتبطان ارتباطًا وثيقًا، مما يعني وجود تكرار الجزئي للسمات الهندسية المقاسة (لن نكون بحاجة إلا لأحد السمتين). وهنا يحدث تحليل المكونات الرئيسية تغييرًا في نظام الإحداثيات من المحاور الأصلية (العرض، الارتفاع) (الموضحة باللون الأسود)، إلى مجموعة جديدة من المحاور (الموضحة باللون الأزرق)، التي تُمثل المحاور الرئيسية لمصفوفة تباينات السمات المقاسة.

باختصار، يُشير المحور الجديد الأول (السهم الأزرق) إلى أقصى تباين في البيانات، بينما يُشير المحور الجديد الثاني (السهم الأزرق المتقطع) إلى أدنى تباين للبيانات. ويُشار إلى تغيير معلم الإحداثيات (من المعلم الأصلي الأسود إلى معلم المركبات الأزرق): بدوران بزاوية V حول مركز سحابة النقاط التي تُمثل البيانات. بإسقاط البيانات على المحاور التي تُطابق أقصى تباين بين السمات، وهو المحور الأول في هذا المثال، يُمكن تقليل أبعاد مساحة البيانات، مع الحفاظ على معظم تباينها.

**7. فوائد طريقة التحليل بالمركبات الرئيسية**

غالبا ما يتم جمع البيانات حول متغيرات كثيرة العدد ومترابطة فيما بينها، وهذا يجعل تفسيرها والكشف عن هيكلها صعبا جدا. وهنا فإن تحويل المتغيرات الأصلية إلى عدد أقل من المتغيرات المستقلة عن طريق تحليل المركبات الرئيسية ACP، يجعل هاتين المهتمين (التفسير والكشف عن هيكلة البيانات) أسهل، ومنه تظهر فوائد طريقة تحليل المركبات الرئيسية في النقاط التالية:

- تخفيض عدد المتغيرات الى عدد اقل من المتغيرات التمثيلية المستقلة، وبالتالي تعديل سحابة النقط بشكل أفضل في فضاء شعاعي ذي أبعاد قليلة (2 أو 3)، والتي تمثل سحابة النقط الأصلية، معالاحتفاظ بأقصى تباين كلي مفسر.
- تصنيف المتغيرات والأفراد، فبالإضافة إلى تقليل الأبعاد للمساحة الأصلية للمتغيرات، فإنه يمكن أيضا استخدام ACP كتقنية تصنيف، وبالتالي يمكن تسليط الضوء على العلاقات بين المتغيرات والأفراد.
- تسمح ACP بتحليل البيانات التي تتعلق بمتغيرات غير المتجانسة (وحدات قياس مختلفة)، من خلال تقديم خيار التحليل باستخدام مصفوفة التغاير (التباينات المشتركة) أو مصفوفة الارتباطات.

**8. الفرق بين تحليل المركبات الرئيسية والتحليل العاملي**

يتشابه تحليل المكونات الأساسية والتحليل العاملي، لأن كلا الإجرائين يستخدمان لتبسيط هيكل مجموعة من المتغيرات، ومع ذلك فإن التحليلين يختلفان في عدة جوانب مهمة أبرزها:

- في تحليل المركبات الرئيسة يتم تقليل واختزال المتغيرات، وانتاج مكونات ذات خصائص إحصائية مثالية، أما في التحليل العاملي فيتم تحديد المتغيرات الكامنة التي مصدر القيم المشاهدة لمتغيرات التنيجة.
- يُعطي تحليل المركبات الرئيسية الأولوية للاحتفاظ بقدر أقصى من المعلومات القصوى، على عكس التحليل العاملي الذي يسعى لإنتاج عوامل قابلة للتفسير والتطبيق في الواقع.
- في تحليل المركبات الأساسية يتم حساب الأبعاد كتركيب خطي لمجموعة من المتغيرات الأصلية، أما في التحليل العاملي فيتم تعريف المتغيرات الأصلية على أنها مجموعات خطية من العوامل.
- في تحليل المركبات الأساسية يكون الهدف حساب أكبر قدر من التباين الكلي في المتغيرات قدر الإمكان، أما الهدف في التحليل العاملي فهو شرح التباين أو الارتباط بين المتغيرات.
- يتم استخدام تحليل المركبات الأساسية لتقليل البيانات إلى عدد أقل من المكونات، ويتم استخدام التحليل العاملي لفهم البنيات الأساسية للبيانات.

**9. أنواع تحليل المركبات الرئيسية**

**أ. الطريقة غير المعيارية (ACP – Non normé):** تستعمل في حالة تجانس المتغيرات، أي أن تكون كل متغيرات جدول البيانات لها نفس وحدة القياس، وفي هذه الحالة يتم طرح متوسط المتغير (العمود) من كل عناصر ذلك العمود للحصول على مصفوفة ممركزة، ولهذا تعرف بـ: المركزة أو التوسيط Centrage.

**ب. الطريقة المعيارية (ACP - Normé):** تستعمل في حالة عدم تجانس المتغيرات، أي أن تكون متغيرات جدول البيانات ليس لها نفس وحدة القياس، وفي هذه الحالة يتم طريق طرح متوسط المتغير (العمود) والقسمة على الإنحراف المعياري لذلك العمود من كل خانة مقابلة في جدول البينات، وهذا حتى يتم الحصول على مصفوفة بيانات ممركزة ومعيارية، ولهذا تعرف بـ: التخفيض Réduction لقسمتها على الانحراف المعياري، أو التوحيد Normalisation لجعلها كل المتغيرات بدون وحدة.

**ملاحظة (1):** تقتصر طريقة ACP غير المعيارية على عملية المركزة فقط، أما طريقة ACP المعيارية فتتعدى إلى عملية التخفيض، وبالتالي تشمل العمليتين معا (المركزة + التخفيض).

**ملاحظة (2):** في حالة البيانات الممركزة فقط (ACP centrée) يتم الاعتماد على مصفوفة الارتباطات بين المتغيرات لاستخراج المركبات أو المكونات الرئيسية، ويتم الاعتماد على مصفوفة التغاير (التباينات المشتركة) في حالة البيانات المعيارية (ACP Normé) لاستخراج المركبات أو المكونات الرئيسية.

**10. شروط طريقة المركبات الرئيسية**

قبل القيام بعملية التحليل البيانات باستخدام المركبات الرئيسية، يجب توفر (5) شروط في البيانات حتى تكون النتائج صحيحة:

- **الشرط الأول:** يجب أن تكون المتغيرات قياسية (كمية) مستمرة، على الرغم من استخدام المتغيرات الترتيبية كثيرا في كثير من الدراسات في العالم الحقيقي.
- **الشرط الثاني:** يجب أن تكون هناك علاقة خطية بين جميع المتغيرات. والسبب أن التحليل بالمركبات الرئيسية يعتمد على معاملات ارتباط بيرسون، لذا يجب أن تكون هناك علاقة خطية بين المتغيرات.
- **الشرط الثالث:** يجب أن يكون هناك كفاية في حجم العينة، وبشكل عام توصي المراجع بحد أدنى من 150 مفردة، أو من 5 إلى 10 مفردات لكل متغير كحد أدنى لحجم العينة، ومن بين اختبارات كفاية العينة مقياس كايزر-ماير- أولكين (KMO) لكفاية أخذ العينات لمجموعة البيانات الإجمالية؛ ومقياس KMO لكل متغير فردي.
- **الشرط الرابع:** يجب أن يكون هناك ارتباطات كافية بين المتغيرات من أجل تقليلها إلى عدد أقل من المكونات. والطريقة التي يستخدمتها SPSS لاكتشاف ذلك هو اختباربارتليب للتكورية Bartlett's Test of sphericity .
- **الشرط الخامس:** يجب ألا يكون هناك قيم متطرفة كبيرة. لأنه سيكون لها تأثير سيئ على نتائج التحليل، وهنا يمكن الاعتماد على اختبار مسافة Mahalanobis في برنامج SPSS.

**11. منهجية التحليل بالمركبات الرئيسية: الخطوات والأدوات**

**أ. إنشاء جدول ومصفوفة البيانات:**

حيث تتكون الصفوف من مفردات العينة (عملاء، منتجات، سنوات، مناطق ...)، وتتكون الأعمدة من المتغيرات المدروسة (كمية)، وهكذا تتشكل مصفوفة $X_{n\times p}= (x_{ij}) : i= 1 \dots n ;\ j= 1 \dots p$، حيث n حجم العينة، وp عدد المتغيرات. و$x_{ij}$ قيمة المتغير $X_j$ للمفردة i.

**ب. تحويل البيانات الأصلية إلى بيانات ممركزة أو معيارية:**

تحسب البيانات الممركزة بطرح متوسط كل متغير (عمود) من كل قيم ذلك العمود، وتحسب البينات المعيارية بطرح متوسط كل متغير (عمود) ثم القسمة على الإنحراف المعياري لذلك العمود، مما يعني تحقيق تجانس البيانات عبر إلغاء مشكلة تضخم البيانات واختلاف وحدات قياس المتغيرات.

**ج. حساب مصفوفة التباينات المشتركة V أو مصفوفة الارتباطات R**

يمكن القيام بتحليل المركبات الرئيسية إما بالاعتماد على مصفوفة التباينات المشتركة أو مصفوفة الارتباطات للمتغيرات، وإن نوع المصفوفة المفضل يعتمد على طبيعة المتغيرات، فيتم الاعتماد على مصفوفة التباينات المشتركة إذا كان للمتغيرات نفس وحدات القياس (تجانس الوحدات)، وتسمي في هذه الحالة التحليل بالمركبات الأساسية غير المعيارية ACP Non normé، ويتم الاعتماد على مصفوفة الارتباطات إذا لم تكن للمتغيرات نفس وحدات القياس (عدم تجانس الوحدات)، ونسمي في هذه الحالة التحليل بالمركبات الرئيسية المعياري ACP Normé.

**د. حساب الأشعة الذاتية والقيم الذاتية** لمصفوفة التباينات المشتركة أو الإرتباطات، هذه المفاهيم المتجذرة في الجبر الخطي، ضرورية في تحديد المكونات الرئيسية. والأشعة الذاتية تمثل الاتجاهات التي تتباين فيها البيانات بشكل أكبر، أما القيم الذاتية فتمثل حجم التباين في اتجاه كل متجه ذاتي. فمثلا إذا كان لدينا مجموعة بيانات ثنائية الأبعاد تحتوي على المتغيرين $X_1$ و $X_2$، الشعاع الذاتي الذي له أكبر قيمة ذاتية يتوافق مع اتجاه أكبر تباين (المكون الرئيسي الأول)، بينما يكون الشعاع الذاتي الثاني متعامدًا على الأول ويمثل ثاني أكبر تباين.

**هـ. تحديد عدد المركبات (المحاور) التي تؤخذ في التحليل**

هناك عدة معايير تسمح بتحديد عدد المركبات (المحاور) التي تؤخذ في التحليل،لأنه لا يؤخذ منها إلا التي تساهم بنسبة كبيرة في التباين الكلي، ومن بين هذه المعايير:

- **معيار كيزر Kaiser:** نأخذ كل القيم الذاتية الأكبر من 1.
- **معيار نسبة المستوي العاملي الأول:** إذا كانت قيمة النسبة: $(\lambda_1)/\sum\lambda_j *100$ أكبر من 80%، ففقدان المعلومات صغير، فلا داعي لإنشاء المستوي العاملي الثاني.

- **معيار نسبة المحور العاملي الثاني:** إذا كانت قيمة النسبة: $(\lambda_2)/\sum\lambda_j *100$ أكبر من 15%، فلا بد من إنشاء المستوي العاملي الثاني.
- **معيار التمثيل البياني للقيم الذاتية:** بعد إنشاء التمثيل البياني لأعمدة للقيم الذاتية نقوم برسم مستقيم شبه أفق يربط بين أكبر عدد من القيم الذاتية، والقيم الكبيرة التي لا تنتمي للمستقيم تمثل المحاور التي تؤخذ في التحليل. أو نربط بين رؤوس الأعمدة بخطوط مستقيمة، وأين تشكل لنا مرفق (Coude)، فنتوقف عند تلك القيمة الذاتية (المرفق يؤخذ).

**و. تشكيل المصفوفة المميزة**

هي مصفوفة تتكون من الأشعة الذاتية المختارة في الخطوة، والتي تمثل المكونات الرئيسية التي سيتم الاحتفاظ بها. فإذا كان لدينا ثلاثة أشعة ذاتية ولكن نقرر الاحتفاظ باثنين فقط بناءً على قيمهم الذاتية، فإن المصفوفة المميزة ستحتوي على عمودين، يمثل كل منهما أحد المتجهين الذاتيين المختارين.

**ز. إعادة تشكيل البيانات على طول المكونات الرئيسية**

الخطوة الأخيرة هي إعادة تشكيل البيانات على طول المكونات الرئيسية. يتم تحقيق ذلك من خلال ضرب مصفوفة البيانات الأصلية في المتجه المميز. النتيجة هي تمثيل جديد للبيانات بناءً على المكونات الرئيسية المختارة.