تحليل المركّبات الأساسية ACP المعيارية

التحليل العاملي AFC

تعاني طريقة المركبات الأساسية البسيطة من مجموعة من المشاكل لعل أهمها:

- إهمال تأثير الوزن النسبي للمتغيرات وحتى للأفراد ولكن هذه المتغيرات القديمة قد تكون غير متجانسة (من حيث السن، الوزن، الحجم،...) فأي معنى يمكن إعطاؤه للمركبات الأساسية التي هي توليفات خطية للمتغيرات غير المتجانسة.
- التحليل غير مرن تجاه التغيير في موقع الأفراد مما قد يغير تماما نتائج التحليل.
- متغير يساهم أكثر في المحاور الأولى إذاكان تباينه أكبر.

1\ استخدامات التحليل العاملي

- ✓ 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 معين، فنبحث هل هذه العناصر تقع في عوامل متميزة تمثل تلك الكليات المحددة. (اختبارات الأداء للكفاءة الأكاديمية).
- ✓ 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 أعد في ألمانيا وآخر في الجزائر، لتعديل الاختبار أو المقياس نلجأ إلى التحليل العاملي).

2\ شروط إجراء التحليل العاملي أو تحليل المركبات الأساسية (FA, PCA)

1- يجب أن تكون جل معاملات الارتباط في المصفوفة أكبر من 0.30. ويمكن **حذف المتغيرات** التي يكون معامل ارتباطها **بأغلب أو كل** المتغيرات الأخرى **ضعيفا أو معدوما**.

- **يجب أن تكون جل معاملات الارتباط في المصفوفة أقل من 0.80.**

2- **يجب أن تكون القيمة المطلقة لمحدد مصفوفة معاملات الارتباط $|\mathbf{R}| \neq 0$** (عمليا يجب أن يكون أكبر من 0.00001). وذلك لتفادي الوقوع في ارتباط خطي بين الأسطر أو الأعمدة، وإن حصل ذلك فغالبا ما نكون أمام مشكلة ارتباط متعدد بين متغيرات النموذج.

3- **يجب أن تكون مصفوفة معاملات الارتباط تختلف عن مصفوفة الوحدة $\mathbf{R} \neq \mathbf{I}$**، وهو ما يعني أن مصفوفة معاملات الارتباط تحتوي على الحد الأدنى من المعلومات لكي تكون قابلة للخضوع للتحليل العاملي،

وهنا نستخدم اختبار Bartlett's test of sphericity:

العينة مختارة عشوائيا من مجتمع مصفوفة معاملات الارتباط به تساوي مصفوفة الوحدة... $\mathbf{H_0}: \mathbf{R} = \mathbf{I}$

إذا كنت تستخدم PCA يمكنك **حذف المتغيرات** التي لا ترتبط بغيرها.

إذا كنت تستخدم PCA من أجل تقليص عدد المتغيرات، يمكنك دائما **إضافة المتغيرات غير** المرتبطة **مع غيرها** قبل تقدير **PCA**. وإلا يمكنك أن **تبحث عن متغيرات أخرى مرتبطة** مع هذا **المتغير الفريد** (جمع بيانات إضافية).

من أجل تحديد المتغيرات الفريدة يمكن تقدير كل متغير على المتغيرات الأخرى، وحساب R^2، **حيث أصغر R^2 تعود** للمتغيرات الفريدة. حيث إذا كنت في **التحليل المحاور الأساسية العاملي** نجدها تحت عنوان جدول Communalities (Initial).

4- **اختبار Kaiser's Measure of Sampling Adequacy**

يستخدم اختبار (MSA) من أجل اختبار مدى وجود متغيرين يتشاركان التباين (Cov) (معامل الارتباط الجزئي) بينهما دون بقية المتغيرات، مما يعني تكرار الجزء المشترك من التباين في التحليل، وهو ما يؤثر سلبا على مصداقية نتائج التحليل. وقيمته تكون مع اختبار **بارتليت** (وأحيانا أسفل جدول الارتباط الجزئي).

$$MSA = \frac{\sum r_{ij}^2}{\sum r_{ij}^2 + \sum pr_{ij}^2}$$

≥ 0.9: رائع، ≥ 0.8: جيد. ≥ 0.7 مقبول. ≥ 0.6 متوسط، ≥ 0.5 بائسة. وأدنى من 0.5 غير مقبول.

3\ مراحل التحليل العاملي

لتجاوز مآخذ طريقة ACP البسيط نبحث عن طريقة لتعيير المتغير وتحويله إلى متغير بدون أبعاد، أكثر الطرق شيوعا هي قسمة المتغيرات على انحرافها المعياري فنحصل على المتغيرات المختزلة (فيستبعد أثر تباين المتغير)، والعودة للقيام بنفس مراحل تحليل ACP البسيط، لكن هناك بعض الاختلافات نحصرها فيما يلي:

أ\ المسافة الإقليدية المعيارية أو الجديدة: نختار مسافة إقليدية جديدة بين نقطتين تناسب التعريف الجديد للمتغير، بحيث تسمح هذه العلاقة بإعطاء الوزن الحقيقي لكل متغير:

$$d^2(u_i, u_{i'}) = \sum_{j=1}^{p} \frac{1}{\sigma_j^2}(x_{ij} - x_{i'j})^2$$

ب\ جدول البيانات المعيارية أو المختزلة **Xcr**:

$$\mathbf{X_{cr} = X_c \ (D_\Sigma)^{-1/2}}$$

$$(\mathbf{D_\Sigma})^{-\frac{1}{2}} = \begin{bmatrix} \frac{1}{\sigma_1} & 0 & \cdots & \cdots & 0 \\ 0 & \ddots & \ddots & & \vdots \\ \vdots & \ddots & \frac{1}{\sigma_j} & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & \frac{1}{\sigma_p} \end{bmatrix}$$

ج\ المصفوفة المستخدمة في التحليل: المصفوفة **R**: في هذه الحالة للبحث عن المحاور لا نستخدم مصفوفة Σ وإنما نستخدم مصفوفة الارتباط بين المتغيرات الابتدائية R:

$$R = \frac{1}{n}\, {}^tX_{cr}\, X_{cr} \Rightarrow R = \frac{1}{n}\, {}^t(X_c\ (D_\Sigma)^{-1/2})\ (X_c\ (D_\Sigma)^{-1/2})$$

$$\Rightarrow R = \frac{1}{n}(D_\Sigma)^{-1/2}\ {}^tX_c\ \ X_c\ (D_\Sigma)^{-1/2}$$

$$\Rightarrow R = (D_\Sigma)^{-1/2}\ \Sigma\ (D_\Sigma)^{-1/2}$$

ملاحظة: يمكن البرهان على أن: $\mathbf{\Sigma_{cr} = \frac{1}{n}\, {}^tX_{cr}\, X_{cr} = R}$

$$R = \begin{pmatrix} Var_1/\sigma_1^2 & Cov_{12}/\sigma_1\sigma_2 & \cdots & Cov_{1p}/\sigma_1\sigma_p \\ Cov_{12}/\sigma_1\sigma_2 & Var_2/\sigma_2^2 & \cdots & Cov_{2p}/\sigma_2\sigma_p \\ \vdots & \vdots & & \vdots \\ Cov_{1p}/\sigma_1\sigma_p & Cov_{2p}/\sigma_2\sigma_p & \cdots & Var_p/\sigma_p^2 \end{pmatrix}$$

د\ معيار تحديد عدد المحاور المستخدمة للتحليل: معيار المساهمة المطلقة للمحور (λ_j):

من الاختلافات بين التحليل البسيط والمعياري هو أن قيم القطر الرئيسي لمصفوفة التحليل أصبحت مساوية ل 1 فقط، فأصبح I_G يساوي عدد متغيرات النموذج (أثر R).

$$Tr(R)= P \Rightarrow I_G = P$$
$$\Rightarrow \sum\lambda_j = P$$
$$\Rightarrow I_G/ P = 1$$

لاختيار عدد المحاور التي تمثل عليها سحابة الأفراد في ACP المعياري، نأخذ فقط المحاور التي عزم العطالة المحمول عليها (أي الجذر المميز **λ_j**) أكبر من **1** **(مفسر)**.

من خصائص ACP المعياري أن تمثيل المتغيرات بواسطة دائرة الارتباط المتعلقة بتمثيل المتغيرات على R^n تماثل بناء –وفق نفس الخطوات– دائرة الارتباط المتعلقة بتمثيل الأفراد على R^p.

تمـرين 1: في المثال السـابق (**1**) الخـاص بالـدخل ومخزون رأس المـال، 1– احسـب المسـافات الإقليديـة المعيارية.

2– احسب مصفوفة معاملات الارتباط ثم استنتج قيمة عزم العطالة الكلي I_G.

$$\mathbf{R} = \begin{pmatrix} \mathbf{1} & \mathbf{0.933} & \mathbf{0.942} \\ \mathbf{0.933} & \mathbf{1} & \mathbf{0.906} \\ \mathbf{0.942} & \mathbf{0.906} & \mathbf{1} \end{pmatrix} \Rightarrow I_G = tr(R) = 3$$

تمرين 2: تعطى المصفوفة $\sum$ فقط حيث:

$$\mathbf{R} = \begin{pmatrix} \dots & \mathbf{0.92} & \dots \\ \dots & \dots & \mathbf{0.18} \\ \mathbf{-0.6} & \dots & \mathbf{1} \end{pmatrix} \Leftarrow \sum = \begin{pmatrix} \mathbf{15} & \mathbf{16} & \dots \\ \dots & \mathbf{20} & \mathbf{4} \\ \mathbf{-12} & \dots & \mathbf{25} \end{pmatrix}$$

1– أكمل المصفوفة $\sum$، واحسب قيم المصفوفة R ووضح بماذا تسميان.

2– احسب عزم العطالة الكلي I_G باستخدام طريقة ACP البسيط وطريقة AFC.

3– عدد المحاور المنسبة للتحليل، إذا علمت أن المساهمة النسبية لعزم العطالة الكلي I_G هي على التوالي 50% و 35% بالنسبة لكل طريقة.

4– وضح أكثر المتغيرات القديمة والجديدة تفسيرا لعزم العطالة الكلي I_G.

في المثال المرفق: استخدام ACP المعيـاري يلغي الأثر المهيمن للمتغيرين DET و DEF اللذين يمثلان في المتوسط 19% و 30% على التوالي بالنسبة إلى البنود الأخرى.

$$I_G = \sum\lambda_j = P = 1 \qquad (37)$$

المحاور الثلاثـة الأولى تحقـق شرط تجاوز عزم العطالة 1، لكنهـا تفسر مجتمعة 75.6%، لـذلك يفضل التوسع إلى غاية المحور الخامس بحيث تفسر المحاور الخمسة الأولى مجتمعة 91% من I_G.

(ص 37)، وتبقى مفاضلة المعايير المحددة لعدد المحاور المناسبة للتحليل اختيارية، وخاضعة لمزيد من الدراسة والتحليل. – لا يختلف تحليل الشكل (ص 41) عنه في تحليل ACP المعياري، فالمحور 2 غير الاتجاه، الإسقاط على الفضاء الجزئي (1-2-3) (ص 42): يظهر ارتفاع 1968 عن بقية السنوات على المحور Δ_3.

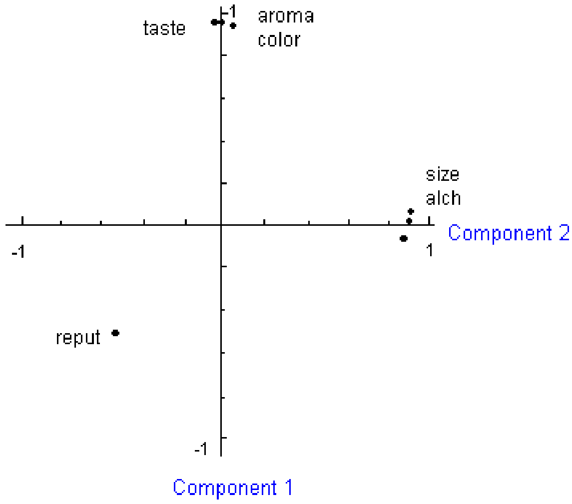
علـــى دائـــرة الارتبـــاط (ص 43) تبـــدو مجموعـــة متجانســـة مـــن المتغـــيرات {LOG, CMI, AGR, ACS, EDU} في تعـاكس مـع المتغير DET، بينما نلاحـظ أن المتغير DEF سيء التمثيل مقارنة بتمثيله الجيد في حالة تحليل ACP البسيط، ويعزى ذلك إلى إلغاء أثره المهيمن بفعل التحليل المعياري، مما جعل مساهمته في المستوي الأول تصبح ضعيفة.

4\ التدوير: التشبعات والمنحنى توضحان انقسام المتغيرات إلى مجموعتين تتناسبين مع المحورين، فمن الأفضل تدوير المعلم ليصبح المحورين يمران عبر أو قرب مجموعة المتغيرات التي تمثلهما.

Rotated Component Matrix[a]

	Component	
	1	2
TASTE	.960	-.028
AROMA	.958	1.E-02
COLOR	.952	6.E-02
SIZE	7.E-02	.947
ALCOHOL	2.E-02	.942
COST	-.061	.916
REPUTAT	-.512	-.533

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.
a. Rotation converged in 3 iterations.



أ\ عدد المركبات في حل التدوير:

أولا بعد الحصول على مصفوفة التشبعات **قبل التدوير**، حددنا عدد المركبات المستخرجة ونتائج هذه الحالة سنأخذها كمرجع، ثم نقارنها مع نتائج **حالة التدوير**: فإذا حدثت تحولات فعلية في النتائج (أصبح محور ذو تشبعات قوية مع مجموعة المتغيرات بعد أن كانت ضعيفة مع كل أو جل المتغيرات) في هذه الحالة نتبنى النتيجة الجديدة، وإلا نحافظ على النتائج الأولية ونرفض الحل الجديد.

استخراج عدد المركبات في البداية قد يكون إما مفرطا أو مقلا، وكلاهما حالة سيئة. لذا نجد أن (Wood, Tataryn, and Gorsuch (1996, *Psychological Methods*, *1*, 354-365)

قد درسا أثر الاستخراج المبالغ فيه والمقل في التحليل العاملي باستخدام طريقة تدوير الفاريماكس rotation varimax. فخلصا إلى الاستخراج المبالغ فيه overextraction له آثار أقل من الاستخراج المقل underextraction على تركيب العوامل وتشبعاتها. وكانت الحالة الأولى (المبالغ) سببا في مشكلة كبيرة في حالة وجود عامل حقيقي واحد، وتم استخراج (مبالغ فيه) عاملين، وعدم وجود متغيرات فريدة (التي لا تشارك التباين مع غيرها). في هذه الحالة وجد Wood أن التشبعات للعامل الأول غير المدور تكون مشابهة لها للعامل الحقيقي، مع تشبعات صغيرة للعامل الثاني، في حين تنعكس الصورة بعد التدوير، فيقلل التدوير بشكل كبير من قيم التشبعات للعامل الأول وتنخض بشكل كبير للعامل الثاني.

ملاحظة: يقترح الباحث في هذه الحالة إضافة **عامل أو متغير فريد** (**إضافي**) بأرقام عشوائية ويدخل يدويا في **مصفوفة الارتباط**، بحيث يكون **معامل ارتباطه 0** مع **بقية المتغيرات**. **وبعد تحديد عدد المتغيرات يحذف** هذا **المتغير** أو العامل.

والأفضل القدرة على الوصول إلى الاستخراج الصحيح.

ب\ التباين المفسر بعد التدوير من طرف المركبات المستخرجة:

ويساوي إلى مجموع مربعات التشبعات (SSL) sum of squared loadings. والذي يشبه هنا الجذر المميز. أحيانا يكون المركب له جذر مميز موجب، وبعد التدوير يتبين أن SSL أقل من 1.

هام: يجب الانتباه للتشبعات بعد التدوير، فإن زاد تشبع أحد المتغيرات فقط بشكل كبير لعامل واحد من العوامل، فإن هذا العامل لا يعتبر محدد بشكل جيد.

إذا زاد تشبع متغيرين بشكل كبير لعامل من العوامل، يتم اختبار مدى وجود ارتباط بينهما وليس مع البقية.

ج\ تسمية المركبات المستخرجة:

1- يجب الانتباه إلى التشبعات القوية أي معاملات ارتباط المركبات مع المتغيرات القديمة.

2- يجب الانتباه إلى إشارة معامل الارتباط. لأن فهم العلاقات القوية الطردية والعكسية، بالإضافة إلى فهم المتغيرات التي لا يتأثر بها العامل أو المركبة، يجعل التسمية أسهل وأدق.

د\ مربعات التشبعات SSL:أو Communalities:

وتقدم SSL مساهمة تباين المتغيرات القديمة في المركبات الجديدة بعد التدوير.

Communalities

	Initial	Extraction
COST	1.000	.842
SIZE	1.000	.901
ALCOHOL	1.000	.889
REPUTAT	1.000	.546
COLOR	1.000	.910
AROMA	1.000	.918
TASTE	1.000	.922

Extraction Method: Principal Component Analysis.

هـ\ التدوير العمودي مقابل التدوير المائل:

1- التدوير العمودي: بما أن التدوير عمودي فإن المركبات مستقلة إحصائيا وبالتالي فهي غير مرتبطة فيما بينها. **أ- طريقة التدوير فاريماكس VARIMAX:** الأشهر والأكثر استخداما، هدفها الأساسي التقليل من التداخل بين المركبات، كما تعمل على زيادة التشبعات المرتفعة وتقليل التشبعات المنخفضة داخل كل مركبة.

ب- طريقة التدوير QUARTIMAX: تعمل على زيادة التشبعات المرتفعة وتقليل التشبعات المنخفضة داخل كل متغير.

ج- طريقة التدوير EQUAMAX: تعمل على تبسيط كل من المركبات والمتغيرات.

2- التدوير المائل: طريقة التدوير QUARTIMAX:

هذه الطرق لا تنتج مركبات مستقلة (غير مرتبطة)، ولكنها تناسب حالات يكون فيها توزيع مجموعات المتغيرات بشكل يجعل من التدوير المائل أفضل من التدوير القائم، كما يظهر في الشكل الموالي: