وزارة التعليم العالي والبحث العلمي

جامعة عبد الحميد بن باديس-مستغانم

كلية العلوم الاقتصادية، التجارية وعلوم التسيير

قسم العلوم الاقتصادية

مطبوعة علمية بيداغوجية مقدمة لطلبة السنة الثالثة ليسانس

تخصص: اقتصاد كمي

مادة :

تحليل البيانات

من اعداد:

يخلف عبد الله

السنة الجامعية : 2024-2025

توطئة

تُعد مادة تحليل البيانات من الركائز الأساسية في التكوين الأكاديمي لطلبة السنة الثالثة ليسانس تخصص اقتصاد كمي، حيث تمثل حلقة وصل بين النظريات الاقتصادية والتطبيقات العملية، حيث تهدف هذه المطبوعة إلى تزويد الطلبة بالمعارف والأدوات اللازمة لتحليل البيانات الاقتصادية واستخلاص النتائج التي تساعد في اتخاذ القرارات الاقتصادية المستندة إلى أسس علمية.

تتناول المطبوعة المفاهيم الأساسية في تحليل المعطيات الإحصائية، بما في ذلك:

1. الحساب المصفوفي
2. تحليل الارتباط و التباين المشترك
3. التحليل العاملي وتحليل المكونات الرئيسية(ACP)
4. التحليل التناظري
5. الترتيب
6. استخدام الأدوات الإحصائية في التحليل العاملي
7. دراسات حالة

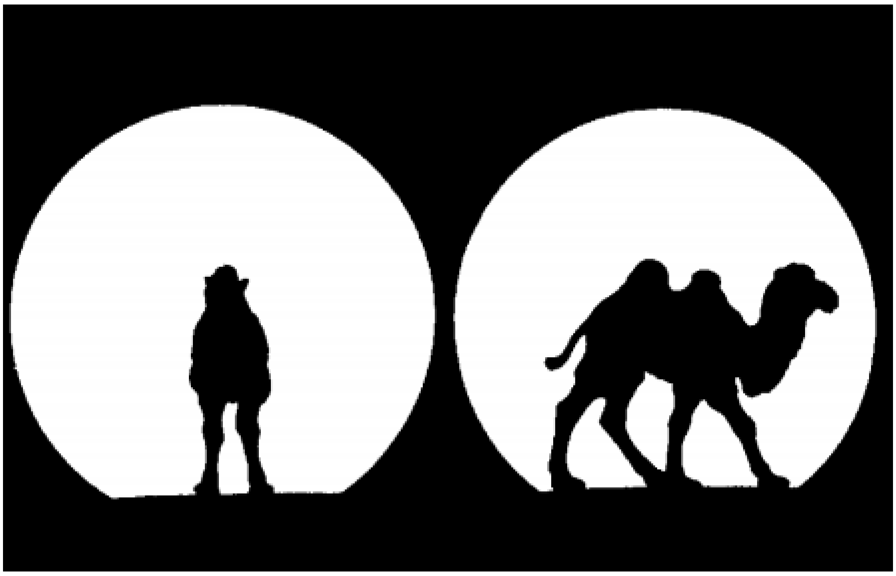
تركز المطبوعة على الجوانب التطبيقية، حيث تتضمن أمثلة عملية وتمارين موجهة تعتمد على برمجيات تحليل البيانات مثل Excelو SPSS و R، مما يُكسب الطالب مهارات عملية تُمكنه من التعامل مع قواعد البيانات الكبيرة ومعالجتها بفعالية.

نسعى من خلال هذه المطبوعة إلى تمكين الطالب من:

1. فهم وتحليل المعطيات الاقتصادية باستخدام الأساليب الكمية.
2. تفسير النتائج الاقتصادية واستخلاص المؤشرات الدالة.
3. توظيف التحليل الإحصائي في صياغة النماذج الاقتصادية والتنبؤات.

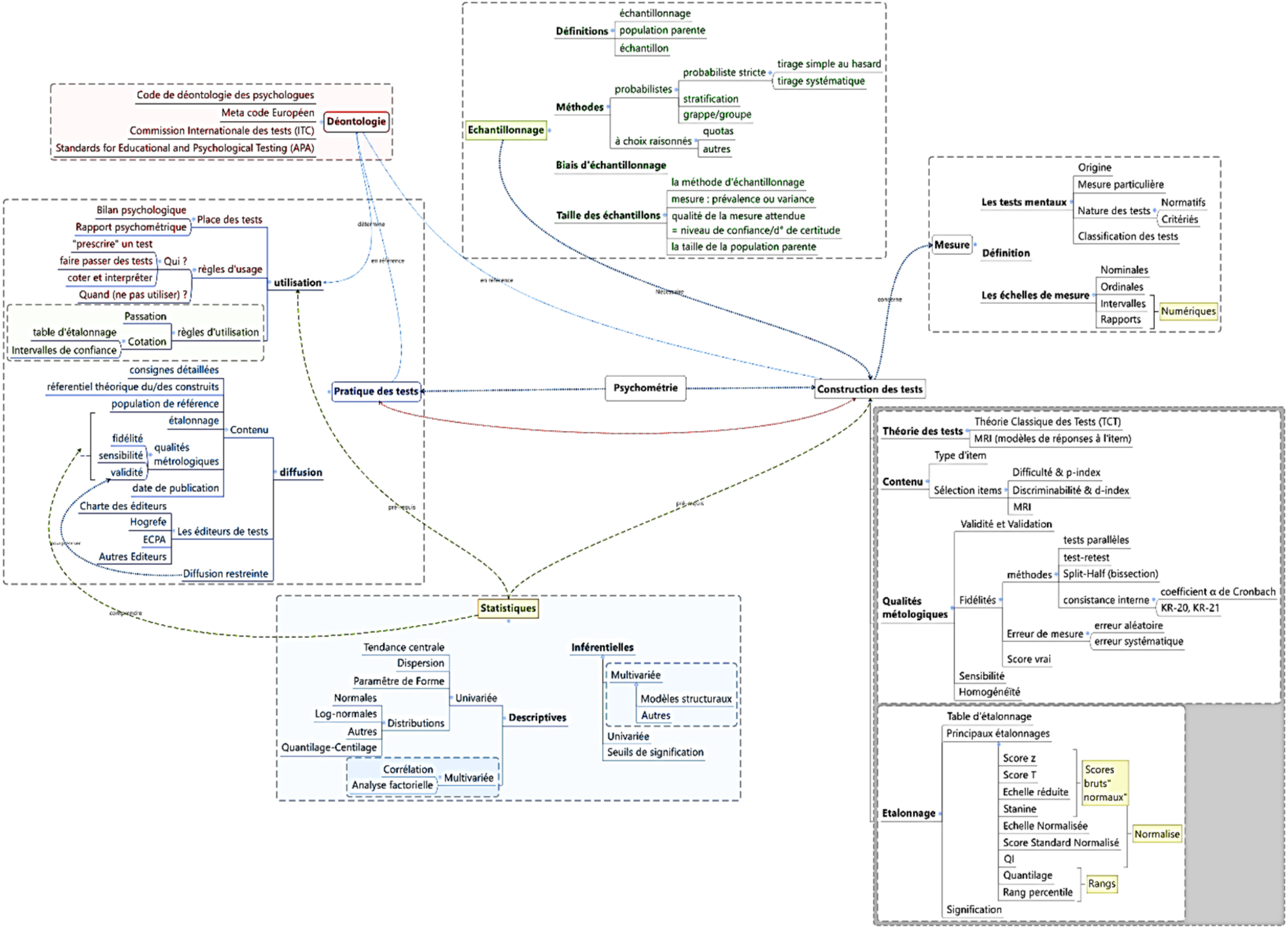
نأمل أن تحقق هذه المطبوعة الأهداف المرجوة وأن تكون مرجعًا علميًا قيمًا يُرافق الطلبة في مسيرتهم الأكاديمية والمهنية.

مع تمنياتنا لكم بالتوفيق والنجاح.



لدينا صورتان بأبعاد مختلفة. عندما يكون هناك أكثر من بُعدين، يصبح من شبه المستحيل التعرف على الحيوان في الشكل، بينما في بعدين فقط، يكون من السهل تصور الشكل ذهنيًا وتحديد الحيوان بسهولة.

يهدف التحليل العاملي إلى إيجاد فضاءات ذات أبعاد أصغر تقلل من هذه التشوهات إلى الحد الأدنى.



Déontologie
Code de déontologie des psychologues
Meta code Européen
Commission Internationale des tests (ITC)
Standards for Educational and Psychological Testing (APA)
Echantillonnage
Définitions
échantillonnage
population parente
échantillon
Méthodes
probabilistes
probabiliste stricte
tirage simple au hasard
tirage systématique
stratification
grappe/groupe
à choix raisonnés
quotas
autres
Biais d'échantillonnage
Taille des échantillons
la méthode d'échantillonnage
mesure : prévalence ou variance
qualité de la mesure attendue
= niveau de confiance/d° de certitude
la taille de la population parente
Mesure
Les tests mentaux
Origine
Mesure particulière
Nature des tests
Normatifs
Critériés
Classification des tests
Définition
Les échelles de mesure
Nominales
Ordinales
Intervalles
Rapports
Numériques
utilisation
Place des tests
Bilan psychologique
Rapport psychométrique
règles d'usage
"prescrire" un test
Qui ?
faire passer des tests
coter et interpréter
Quand (ne pas utiliser) ?
règles d'utilisation
Passation
Cotation
table d'étalonnage
Intervalles de confiance
diffusion
Contenu
consignes détaillées
réferentiel théorique du/des construits
population de référence
étalonnage
qualités métrologiques
fidélité
sensibilité
validité
date de publication
Les éditeurs de tests
Charte des éditeurs
Hogrefe
ECPA
Autres Editeurs
Diffusion restreinte
Pratique des tests
Psychométrie
Construction des tests
détermine
en référence
en référence
concerne
pré-requis
pré-requis
comprendre
Théorie des tests
Théorie Classique des Tests (TCT)
MRI (modèles de réponses à l'item)
Contenu
Type d'item
Sélection items
Difficulté & p-index
Discriminabilité & d-index
MRI
Qualités métologiques
Validité et Validation
Fidélités
méthodes
tests parallèles
test-retest
Split-Half (bissection)
consistance interne
coefficient α de Cronbach
KR-20, KR-21
Erreur de mesure
erreur aléatoire
erreur systématique
Score vrai
Sensibilité
Homogénéité
Etalonnage
Table d'étalonnage
Principaux étalonnages
Score z
Score T
Echelle réduite
Stanine
Echelle Normalisée
Score Standard Normalisé
QI
Quantilage
Rang percentile
Scores bruts" normaux"
Normalise
Rangs
Signification
Statistiques
Descriptives
Univariée
Tendance centrale
Dispersion
Paramètre de Forme
Distributions
Normales
Log-normales
Autres
Quantilage-Centilage
Multivariée
Corrélation
Analyse factorielle
Inférentielles
Multivariée
Modèles structuraux
Autres
Univariée
Seuils de signification

مقدمة

يُعد تحليل البيانات او المعطيات من أهم أدوات البحث العلمي في مختلف المجالات، وخاصة في ميدان الاقتصاد الكمي، حيث يعتمد صُنّاع القرار والباحثون على البيانات لفهم الظواهر الاقتصادية، استشراف المستقبل، وحل المشكلات المعقدة.

كما يمثل تحليل المعطيات عملية منهجية تهدف إلى جمع البيانات، معالجتها، وتنظيمها بهدف استخراج المعلومات والمعارف التي تساعد على تفسير الظواهر وتقديم الحلول العلمية. وتشمل هذه العملية استخدام أدوات وأساليب إحصائية وكمية لتحليل العلاقات بين المتغيرات ودراسة أنماطها.

وتتعدد أهمية تحليل المعطيات كأداة إحصائية وصفية، ويمكن ان نذكر:

1. يساعد في تحليل الاتجاهات الاقتصادية وتحديد الأنماط الخفية.
2. يدعم عملية اتخاذ القرارات الاقتصادية والإدارية المبنية على أدلة رقمية.
3. يساهم في تقييم السياسات الاقتصادية واختبار مدى فعاليتها.
4. يُمكّن من التنبؤ بالتغيرات المستقبلية عبر بناء نماذج كمية دقيقة.

ويعد التحليل العاملي (Factor Analysis) من الأساليب الإحصائية متعددة المتغيرات التي تهدف إلى تقليل حجم البيانات وتحديد العوامل الخفية (الكامنة) التي تفسر العلاقات بين المتغيرات. هناك عدة أنواع وأساليب للتحليل العاملي، من أبرزها:

التحليل العاملي الاستكشافي (Exploratory Factor Analysis – EFA) : و الذي يهدف الى الكشف عن الهيكل الأساسي للبيانات وتحديد العوامل الكامنة التي تربط بين المتغيرات، ويستخدم عادة عند عدم توفر فرضيات مسبقة حول العلاقات بين المتغيرات.

وتتعدد الأدوات المستخدمة في تحليل المعطيات، منها برامج مثل:

1. Excel لتحليل البيانات الأساسي.
2. SPSS للتحليل الإحصائي المتقدم.
3. R وPython للتحليل الكمي والمعالجة المتقدمة للبيانات الضخمة.

يُعتبر تحليل المعطيات مهارة أساسية للطلبة والباحثين في تخصص الاقتصاد الكمي، حيث يربط بين النظريات الاقتصادية والواقع العملي. ومن خلال إتقان أساليبه وأدواته، يصبح الطالب قادرًا على معالجة المشكلات الاقتصادية، وتفسير المؤشرات، والتنبؤ بالاتجاهات الاقتصادية، مما يؤهله للمساهمة بفعالية في صنع القرار الاقتصادي المستند إلى أدلة علمية وكمية.

برنامج مادة تحليل البيانات

الفصل الأول: مقدمة إلى تحليل البيانات

تمهيد:

تعتبر تحليل المعطيات او المعطيات عملية تحويل البيانات الخام إلى معلومات ذات معنى تُستخدم لدعم اتخاذ قرارات مبنية على الأدلة، حيث تتضمن هذه العملية مراحل متعددة تبدأ بجمع البيانات وتنظيفها، مرورًا بمعالجتها، وانتهاءً بتفسير النتائج واستخدامها بشكل فعّال في صنع القرار.

كما تكمن أهمية تحليل المعطيات في قدرتها على تمكين الأفراد والمؤسسات من اتخاذ قرارات دقيقة استنادًا إلى رؤى مستخلصة من البيانات، مما يضمن التخطيط الفعّال والتنبؤ بالمستقبل، كما تتيح التعرف على الأنماط والعلاقات بين المتغيرات، مما يسهم في تحسين الكفاءة والإنتاجية وتقليل التكاليف. إضافة إلى ذلك، تدعم تحليل المعطيات الابتكار من خلال تقديم رؤى جديدة لتحسين المنتجات والخدمات، وتساعد على إدارة المخاطر بتحديد التحديات المحتملة ووضع خطط استجابة مناسبة.

أما في الواقع العملي، تستخدم تحليل المعطيات في مختلف المجالات، ففي الأعمال، يتيح فهم سلوك المستهلك وتحسين استراتيجيات التسويق. وفي الصحة، يساهم في تطوير علاجات جديدة وتحسين جودة الرعاية الصحية. أما في التعليم، فيساعد على متابعة أداء الطلاب وتحسين المناهج الدراسية، بينما يسهم في الاقتصاد بالتنبؤ بتوجهات الأسواق ودراسة المؤشرات الاقتصادية لدعم القرارات الاستراتيجية. تحليل المعطيات ليس مجرد عملية تقنية، بل هو أداة أساسية للنجاح في شتى المجالات.

تعريف تحليل المعطيات وأهميته

ترجع أصول التحليل العاملي الاستكشافي (AFE) إلى أعمال **تشارلز سبيرمان*** [1] (1904)، حيث قدم مفهوم "العامل العام" لتفسير العلاقات بين المتغيرات من خلال نموذج الذكاء العام(g factor) ، وكان هدفه الأساسي هو تحليل التباين المشترك بين المتغيرات لتحديد العوامل الكامنة التي تؤثر عليها.

أما مصطلح التحليل العاملي، فقد ظهر لاحقًا مع **لويس ثورستون** (1931)، الذي طور نظرية العوامل المتعددة، معارضًا نموذج سبيرمان الأحادي للعامل العام، واقترح ثورستون أن الذكاء البشري يتكون من عدة قدرات مستقلة، مما أدى إلى تطوير تحليل العوامل المتعددة، وهو حجر الأساس للتحليل العاملي الاستكشافي.

وفي عام 1933، قدم **هارولد هوتيلينغ** مفهوم تحليل المركبات الرئيسية(ACP) ، والذي يهدف إلى تبسيط مجموعة كبيرة من المتغيرات من خلال تقليل عدد الأبعاد، مع الاحتفاظ بأكبر قدر ممكن من المعلومات، ويُعتبر ACP اليوم أحد أكثر الأدوات شيوعًا في تحليل البيانات، خاصة في مجالات مثل الاقتصاد، علم النفس، والذكاء الاصطناعي.

ولقد تطورت الأساليب العاملية وتعددت تطبيقاتها منذ ذلك الحين، حيث شهدت تقنيات التحليل العاملي تطورًا هائلًا، فظهرت العديد من المناهج والأساليب التي تتناسب مع طبيعة البيانات المختلفة، فعلى سبيل المثال:

- التحليل العاملي التوكيدي(CFA) ، الذي يهدف إلى اختبار مدى صحة البنية العاملية المفترضة للبيانات بناءً على نماذج مسبقة.
- التحليل العاملي التناظري (AFC) ، الذي يُستخدم لتحليل الجداول الضخمة التي تحتوي على بيانات فئوية، وقد طوره **ژان بول بن زكري** في الستينيات.(Benzecri, 1982)

تعريف التحليل العاملي : يشير مصطلح التحليل العاملي إلى مجموعة من طرق تحليل البيانات التي غالبًا ما تشمل (وأحيانًا عن طريق الخطأ) كلًا من تحليل المركبات الرئيسية (ACP) والتحليل العاملي الاستكشافي (AFE) بمعناه الإنجليزي. الهدف العام لهذه التقنيات ليس فقط تقليل البيانات) كما في (ACP بل تُعتبر وسيلة لتحديد العوامل المنظمة لسحابة النقاط بغرض تقليل عدد المتغيرات) البحث عن المركبات

* التعريف بالأعلام موجود في الملحق اخر المطبوعة

فيACP ، والبحث عن المتغيرات الكامنة في .(AFE ولتلخيص ذلك، تتيح هذه التقنيات "تكثيف المعلومات الأولية" واكتشاف هياكل تنظيم البيانات (سواء كانت مركبات أو متغيرات كامنة).

ليس الوصول إلى هذه التقنيات التحليلية أمرًا سهلًا دائمًا، حتى وإن ساعد الانفجار في عدد البرامج المخصصة وغير المخصصة على تنفيذها (وأحيانًا يؤدي ذلك إلى استخدامها بشكل غير صحيح). سنعرض هنا تحليل المركبات الرئيسية وتحليل العوامل المشتركة والخاصة، وهما طريقتان استكشافية؛ إذ لا تحددان مسبقًا المتغيرات التي يجب ربطها بعوامل أو مركبات معينة، وإنما تصفان البيانات بالنسبة للسكان الذين جُمعت منهم البيانات.

حاليًا، هناك تطورات هامة في مجال تحليل البيانات، مثل التقنيات المسماة بالتحليل العاملي التأكيدي، والتي تتيح اختبار فرضيات مسبقة تتعلق بعدد العوامل وانتماء كل متغير إلى عامل معين. كما يمكن باستخدام تقنيات أكثر تعقيدًا (مثل نماذج المعادلات الهيكلية) اختبار العلاقات أو نسب السببية المتعددة بين العوامل (المتغيرات الكامنة غير المرصودة). سنقتصر هنا على التطرق بإيجاز إلى التقنيات التأكيدية المصاحبة للتقنيات الاستكشافية، وهي الأدوات الأكثر استخدامًا حاليًا في بناء الاختبارات (اختيار البنود ودراسة الصلاحية).

أهمية التحليل العاملي

نستخدم الأساليب العاملية لدراسة مجموعة بيانات ذات حجم كبير للأسباب التالية:

1. تقليل الأبعاد: تساعدنا هذه الأساليب في تحويل مجموعة كبيرة من المتغيرات إلى عدد أقل من المتغيرات الأساسية (العوامل) مع الحفاظ على أكبر قدر ممكن من المعلومات. هذا يُسهل التعامل مع البيانات وتحليلها.
2. تبسيط التمثيل: عند تقليل عدد الأبعاد، يصبح من الممكن تصور البيانات بشكل أوضح وفهم الهياكل والأنماط الخفية داخلها، مما يساعد في التعرف على العلاقات بين المتغيرات.
3. الكشف عن الهيكل الأساسي: تُستخدم الأساليب العاملية لاكتشاف العوامل الرئيسية التي تفسر أكبر قدر من التباين في البيانات، مما يُسهم في تفسير الظواهر المعقدة بصورة مبسطة.
4. تحسين الدقة والتحليل: من خلال التركيز على المتغيرات الأكثر تأثيراً، يمكن للمحلل تقليل التشويش الناجم عن البيانات غير الضرورية وتحسين دقة النتائج المستخلصة من الدراسة.

ولم تعد تقنيات التحليل العاملي مقتصرة على المقاييس الترتيبية أو الفئوية، بل أصبحت تُستخدم على نطاق واسع في مجالات متعددة مثل:

❖ علم البيانات والذكاء الاصطناعي :لتقليل الأبعاد واكتشاف الأنماط المخفية في البيانات الضخمة.

- ❖ علم النفس وعلم الاجتماع :لدراسة السمات الشخصية والعوامل المؤثرة في السلوك البشري.
- ❖ الاقتصاد وإدارة الأعمال :لتحليل الأسواق وفهم سلوك المستهلكين.

لهذا يُعتبر الفهم العميق لكل من تحليل المركبات الرئيسية (ACP) والتحليل العاملي الاستكشافي (AFE)مفتاحًا للانتقال إلى نماذج أكثر تقدمًا، مثل نماذج المعادلات الهيكلية(SEM) ، التي تتيح دراسة العلاقات السببية المعقدة بين المتغيرات الكامنة.

قبل ان نمر الى تقنيات التحليل العاملي يجب أولا تعريف العنصر الأساسي في الإحصاء و هو البيانات

تعريف البيانات (Data) **:** هي مجموعة من المعلومات أو القيم التي يتم جمعها وتسجيلها بطريقة منظمة بهدف تحليلها واستنتاج النتائج منها. تُستخدم هذه البيانات في العديد من المجالات مثل الاقتصاد، والعلوم، والصحة، والهندسة، والتجارة لاتخاذ قرارات مبنية على الأدلة، أو هي مجموعة من الحقائق أو القيم الخام غير المُنظمة أو غير المُعالجة، والتي تُجمَع عبر الملاحظة، القياس، التسجيل، أو أي وسيلة أخرى، ويمكن أن تأتي في أشكال متنوعة مثل:

1. أرقام (مثل درجات الحرارة، الأسعار).
2. نصوص (مثل التعليقات على وسائل التواصل الاجتماعي).
3. صور أو فيديوهات (مثل الصور الطبية).
4. أصوات (مثل تسجيلات المحادثات).

وتتميز البيانات بعدة خصائص أهمها :

1. **غير مُفسَّرة:** تحتاج إلى معالجة أو تحليل لاستخراج المعنى منها.
2. **خام:** قد تكون غير منظمة أو تحتوي على ضوضاء (أخطاء أو معلومات غير ذات صلة).
3. **محايدة:** لا تحمل قيمة في حد ذاتها دون سياق أو هدف.

وتُستخدم البيانات عادة كمادة خام لاستخلاص المعلومات (Information) ثم المعرفة (Knowledge) عبر التحليل. تعتمد عليها مجالات مثل:

1. الذكاء الاصطناعي والتعلم الآلي.
2. البحوث العلمية واتخاذ القرارات.
3. التسويق الرقمي وتحليل السوق.

أنواع البيانات (نوعية، كمية، زمنية...).

يمكن تصنيف البيانات الإحصائية إلى عدة أنواع وفقًا لطبيعتها واستخدامها، وفيما يلي تفصيل لكل نوع مع أمثلة عملية:

1. **البيانات الكمية (العددية)** - (**Quantitative**) **Numerical Data** هي البيانات التي تمثل أرقامًا وقيمًا عددية قابلة للقياس. وتنقسم إلى نوعين:

أ. **البيانات المتقطعة(Discrete Data)** : هي بيانات عددية تأخذ قيمًا محددة ومنفصلة، وعادة ما تكون أعدادًا صحيحة، ولا يمكن أن تحتوي على قيم عشرية بين الأرقام المتتالية.

أمثلة:

1. عدد الطلاب في الفصل (30 ،. 31، 32...)
2. عدد السيارات في موقف السيارات(5 ، 10، 15.(...
3. عدد أفراد الأسرة (2 ، 3، 4 ...) .
4. عدد المنتجات المباعة يوميًا في متجر.

ب. **البيانات المستمرة(Continuous Data)** : هي بيانات عددية يمكن أن تأخذ أي قيمة داخل مدى معين، بما في ذلك القيم العشرية، وتُقاس بدقة باستخدام أجهزة القياس مثل الميزان أو المسطرة.

أمثلة:

1. الطول (165.5 سم، 170.2 سم...).
2. الوزن (65.4 كغ، 72.1 كغ...).
3. درجة الحرارة (20.3° م، 35.7° م...).
4. المدة الزمنية لقطع مسافة معينة (12.5 ثانية، 13.8 ثانية...).

⇦ **ملاحظة :البيانات المستمرة تكون غير محدودة من حيث القيم، بينما البيانات المتقطعة تكون محدودة بأعداد صحيحة فقط.**

2. **نوعية/ وصفية** Categorical Data (**Qualitative**): هي البيانات التي تُستخدم لتصنيف الأشياء أو الأفراد ضمن مجموعات دون أي قيمة عددية محددة. وتنقسم إلى نوعين:

أ. **البيانات الاسمية(Nominal Data)** : هي بيانات تصنيفية لا تحمل أي ترتيب معين، تُستخدم لتمييز الفئات أو التصنيفات المختلفة دون وجود تسلسل منطقي بينها.

أمثلة:

1. الجنس: (ذكر، أنثى).
2. لون العين: (أزرق، أخضر، بني).
3. نوع السيارة: (تويوتا، مرسيدس، فورد).
4. التخصص الجامعي: (اقتصاد، هندسة، طب).

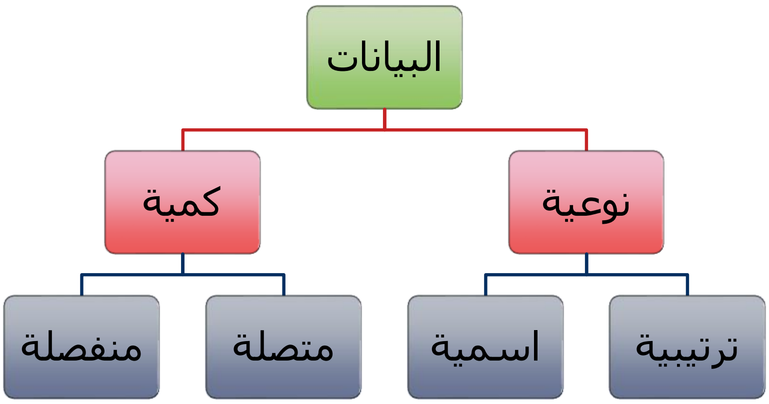
ب. **البيانات الترتيبية**(Ordinal Data) : هي بيانات تصنيفية لكنها تمتلك ترتيبًا معينًا أو مستوى متدرج، لكن الفروقات بين المستويات غير متساوية بالضرورة.

أمثلة:

1. المستوى التعليمي: (ابتدائي، متوسط، ثانوي، جامعي).
2. درجة الرضا: (غير راضٍ، محايد، راضٍ).
3. الترتيب في مسابقة: (الأول، الثاني، الثالث).
4. التصنيف الاقتصادي: (فقير، متوسط، غني).

يمكن ان نلخص أنواع البيانات في الشكل التالي :

شكل رقم (1): تصنيف البيانات



ويمكن المقارنة بين البيانات الكمية و النوعية من خلال الجدول التالي :

الجدول رقم 1: الفرق بين البيانات النوعية والكمية

الخاصية	البيانات الكمية	البيانات النوعية
القيم	أعداد وقيم رقمية	**رموز، أسماء، تصنيفات**
العمليات الحسابية	ممكنة (جمع، طرح، قسمة...)	**غير ممكنة**
أمثلة	الطول، الوزن، الدخل	**لون العين، الجنس، التخصص**

مصادر البيانات :

تُجمع البيانات الإحصائية من مصادر مختلفة حسب طبيعة البحث أو الدراسة، وتنقسم إلى مصدرين رئيسيين:

1. **المصادر الأولية(Primary Sources)**

هي البيانات التي يتم جمعها مباشرة من المصدر الأصلي لأول مرة، دون أن تكون قد تم تسجيلها أو معالجتها من قبل. تُستخدم هذه البيانات في الأبحاث والدراسات التي تتطلب دقة وحداثة.

ومن أهم طرق جمع البيانات الأولية:

أ. **الاستبيانات(Surveys & Questionnaires)** : عبارة عن مجموعة من الأسئلة الموجهة لعينة من الأفراد لجمع معلومات حول موضوع معين، حيث يمكن أن تكون الاستبيانات ورقية أو إلكترونية.

أمثلة:

1. استبيان عن رضا العملاء عن خدمة معينة.
2. استبيان حول تأثير التعليم على فرص العمل.

ب. **المقابلات الشخصية(Interviews)** : يتم جمع البيانات مباشرة من الأفراد عبر طرح أسئلة شفوية،ويمكن أن تكون المقابلات فردية أو جماعية.

أمثلة:

1. مقابلة مع خبراء اقتصاديين حول تأثير التضخم.
2. مقابلة مع الطلاب حول جودة التعليم الجامعي.

ج. الملاحظة المباشرة(Direct Observation) : يتم جمع البيانات عن طريق مراقبة الظواهر أو الأحداث دون تدخل مباشر،وتُستخدم في دراسات السلوك البشري والظواهر الطبيعية.

أمثلة:

1. مراقبة سلوك المستهلكين في المتاجر.
2. دراسة حركة المرور في مدينة معينة.

د. التجارب العلمية(Experiments)

تُستخدم في الدراسات العلمية لاختبار فرضيات محددة، كما يتم التحكم في العوامل المؤثرة وتحليل تأثيرها على الظاهرة المدروسة.

أمثلة:

1. تجربة تأثير دواء جديد على مرض معين.
2. تجربة تأثير الأسمدة على إنتاج المحاصيل الزراعية.

2 . المصادر الثانوية(Secondary Sources)

هي البيانات التي تم جمعها ومعالجتها مسبقًا من قبل جهات أخرى، مثل الحكومات، المؤسسات البحثية، أو المنظمات الدولية. تُستخدم هذه البيانات لتوفير الوقت والتكاليف في الأبحاث.

و تتمثل أهم مصادر البيانات الثانوية فيما يلي :

أ. التقارير الحكومية والإحصاءات الرسمية

تصدرها الجهات الرسمية مثل وزارات التخطيط، التعليم، والصحة، حيث تحتوي على بيانات موثوقة تُستخدم في التحليلات الاقتصادية والاجتماعية.

أمثلة:

1. تقارير الدخل القومي والناتج المحلي الإجمالي.
2. إحصاءات عدد السكان والتوزيع الديموغرافي.
3. تقارير حول معدلات البطالة والتضخم.

ب. الدراسات والأبحاث الأكاديمية

تُنشر من قبل الجامعات والمراكز البحثية حول موضوعات علمية واقتصادية واجتماعية، وتُعدّ مصدرًا غنيًا بالمعلومات المتخصصة.

أمثلة:

1. دراسة عن تأثير الاستثمار في التعليم على النمو الاقتصادي.
2. بحث عن العلاقة بين الصحة والإنتاجية في سوق العمل.

ج. قواعد البيانات والمصادر الإلكترونية: كمواقع الإنترنت، قواعد البيانات المفتوحة، والمصادر الرقمية تحتوي على كميات هائلة من البيانات، وتُستخدم على نطاق واسع في التحليل الإحصائي والاقتصادي.

أمثلة:

1. البنك الدولي :(World Bank) يوفر بيانات عن التنمية الاقتصادية والاجتماعية.
2. منظمة الأمم المتحدة :(UN) توفر إحصاءات حول الفقر، التعليم، والصحة.
3. منظمة الصحة العالمية :(WHO) توفر بيانات حول الأوبئة والأمراض.

د. التقارير الصادرة عن الشركات والمؤسسات الخاصة

تصدرها الشركات الكبرى والمؤسسات المالية لتقييم الأسواق والقطاعات الاقتصادية، حيث تُستخدم في تحليل الاتجاهات الاقتصادية واتخاذ القرارات الاستثمارية.

أمثلة:

1. تقارير البنوك حول مؤشرات الاقتصاد المحلي.
2. بيانات المبيعات من شركات مثل أمازون أو أبل.

والجدول التالي يوضح الفروق الأساسية بين المصادر الأولية و الثانوية عند جمع البيانات:

الجدول رقم 2 :الفرق بين المصادر الأولية والثانوية

المعيار	المصادر الأولية	المصادر الثانوية
طريقة الجمع	تُجمع مباشرة من المصدر	تم جمعها سابقًا
الزمن والتكلفة	مكلفة وتحتاج لوقت	أقل تكلفة وأسرع
دقة البيانات	عالية لأنها غير معالجة	قد تكون أقل دقة لأنها معالجة
أمثلة	الاستبيانات، المقابلات، الملاحظة	التقارير الحكومية، قواعد البيانات

أهمية اختيار مصدر البيانات المناسب

3. **تحديد دقة البيانات**: المصادر الأولية تعطي بيانات أكثر دقة، بينما المصادر الثانوية تكون أكثر شمولًا.
4. **تقليل التكاليف والوقت**: استخدام البيانات الثانوية يقلل من نفقات البحث مقارنةً بجمع بيانات جديدة.
5. **تحقيق أهداف البحث**: يجب اختيار المصدر المناسب حسب نوع الدراسة والأهداف المرجوة.

إذا كنا بحاجة إلى بيانات حديثة ومحددة، فسنستخدم المصادر الأولية.

إذا كنا تبحث عن بيانات جاهزة وموثوقة، فسنستخدم المصادر الثانوية.

المتغير:

هو عنصر قابل للقياس يأخذ قيمًا مختلفة عبر الأفراد أو الزمن، ويُستخدم لتمثيل ظاهرة ما وتحليلها إحصائيًا.[2]

أنواع المتغيّرات في تحلیل البيانات

1) **المتغيرات الكمية[3](Quantitative Variables)** : قيمها أرقام تمثل مقدارًا أو كمية.

- **متواصل(Continuous)** يأخذ أي قيمة داخل مجال معين.
مثال: الدخل، الوزن، معدل النمو.

[2] University of Minnesota – *Principles of Statistics*
https://open.lib.umn.edu/introstat/chapter/1-2-variables-and-types-of-data/

[3] Australian Bureau of Statistics – Statistical Language
https://www.abs.gov.au/statistics/understanding-statistics/statistical-language/quantitative-and-qualitative-data

❖ **متقطع(Discrete)** أعداد صحيحة فقط.

مثال: عدد الطلاب، عدد الحوادث.

2) **المتغيرات النوعية**[4] **(Qualitative / Categorical Variables)** : تمثل **فئات** أو **تصنيفات**.

❖ **اسمي(Nominal)** فئات بدون ترتيب. مثال: الجنس، نوع السيارة.

❖ **ترتيبي(Ordinal)** فئات لها ترتيب منطقي. مثال: مستوى التعليم (ثانوي – جامعي – ماجستير).

المتغيرات في النماذج الإحصائية والاقتصادية

1) **المتغير المستقل(Independent Variable)** هو المتغيّر الذي **يؤثر** في غيره. مثال في الاقتصاد: الإنفاق على التعليم.
2) **المتغير التابع(Dependent Variable)** هو المتغير الذي **يتأثر** بغيره. مثال: معدل النمو الاقتصادي.
3) **متغيرات ضابطة(Control Variables)** : متغيرات تُثبّت أو تُؤخذ بعين الاعتبار لاستبعاد تأثيرها على العلاقة بين المتغيرات المستقلة والتابعة.
4) **متغيرات وسيطة (Mediating) أو مُعدِّلة(Moderating)** : تُستخدم في نماذج متقدمة لفهم كيف أو متى يحدث التأثير.

تمثيل المتغيّرات في تحليل البيانات

يتم تمثيل المتغيّرات في:

- الجداول(Variables = Columns)
- الرسوم البيانية(Boxplot, Histogram...)
- النماذج الإحصائية(OLS, ARDL...)
- الخوارزميات في تعلم الآلة(Features)

خصائص مهمة للمتغيّرات[5]

1) **المجال** :**(Range)** أدنى وأعلى قيمة
2) **القيم المفقودة(Missing values)**
3) **التفرّد والتميّز(Uniqueness)**

[4] UCLA Institute for Digital Research & Education https://stats.oarc.ucla.edu/other/mult-pkg/whatstat/types-of-variables/

[5] DataCamp – Data Types https://www.datacamp.com/tutorial/types-of-data

4) **الانحراف والتوزيع(Distribution)**

التحليل الاحصائي للمتغيرات :

يمكننا عمومًا التمييز بين ثلاثة أنواع من التحليلات الإحصائية: الإحصاءات الوصفية لمتغير واحد، والتحليلات أحادية المتغير (غالبًا ما تُسمى التحليلات الأحادية)، والتحليلات متعددة المتغيرات (والتي يُشار إليها أحيانًا بشكل غير دقيق باسم التحليلات المتعددة).

تُستخدم الإحصاءات الوصفية لمتغير واحد لوصف البيانات، وهي مفيدة لاكتشاف المشكلات المحتملة فيها.

أما التحليلات الأحادية والمتعددة المتغيرات، فتُستخدم للمقارنات الإحصائية (الحصول على قيمة p-value)، ولكن فقط التحليلات متعددة المتغيرات تُمكّن من أخذ عوامل الإرباك في الاعتبار.

التحليل الإحصائي أحادي المتغير (Univariate Analysis)

مفهوم التحليل أحادي المتغير

التحليل أحادي المتغير هو تحليل إحصائي يُستخدم لدراسة وتوصيف سلوك **متغير واحد فقط** في البيانات دون النظر إلى تأثير أي متغيرات أخرى، ويهدف هذا التحليل إلى فهم توزيع البيانات، مقاييس النزعة المركزية، والتشتت، وتمثيل البيانات بيانياً.

طرق التحليل أحادي المتغير

مقاييس النزعة المركزية (Measures of Central Tendency)

مقاييس النزعة المركزية(measures of central tendency هن المقاييس التي تحاول أن تصف نقطة تجمع المشاهدات، وتعود فكرتها إلى الباحث الإنجليزي فرانسيس جالتون. هذه المقاييس هي المتوسط الحسابي والوسيط الحسابي والمنوال، وتُستخدم لتحديد القيمة التي تتركز حولها البيانات، وتشمل:

المتوسط الحسابي(Mean) : يُحسب بجمع جميع القيم وقسمتها على عددها.، **وتكتب علاقته كالتالي:**

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

ومن خواصه:

1. يعتمد على جميع القيم والمشاهدات
2. هو نقطة اتزان المشاهدتان
3. مربع الانحرافات اقل ما يمكن عن الوسط
4. اقل مقاييس النزعة المركزية تأثرا بالتقلبات العينية
5. يتأثر بالقيم المتطرفة والقيم الشاذة لذا لا يصلح للتوزيعات الملتوية
6. لا يصلح في حالة الفئات المفتوحة (لعدم وجود مركز فئة)
7. مجموع انحرافات القيم عن المتوسط الحسابي يساوي الصفر.

مثال: إذا كانت أعمار 5 طلاب هي (20، 22، 21، 19، 23)، فإن:

$$\text{المتوسط} = \frac{20+22+21+19+23}{5} = 21$$

الوسيط(Median) : هو القيمة التي تقع في منتصف البيانات بعد ترتيبها تصاعديًا، ومن خواصه:

1. لا يتأثر بالقيم المتطرفة
2. يستخدم في التوزيعات الملتوية
3. يفضل استخدامه في حالة الفئات المفتوحة
4. يأتي بعد الوسط في تأثره بالتقلبات العينية

مثال: إذا كانت الدرجات (45، 50، 55، 60، 65)، فالوسيط هو 55.

إذا كان عدد القيم زوجيًا، يؤخذ متوسط القيمتين في المنتصف.

المنوال(Mode) : هو القيمة الأكثر تكرارًا في مجموعة البيانات، من خواصه:

1. غير ثابت
2. يتأثر بطول الفئة
3. يفضل عندما يكون المقياس اسمي
4. لا يعتمد عليه في حالة الإحصاءات اللاحقة

مثال: إذا كانت البيانات (2، 3، 3، 4، 5، 3)، فإن المنوال هو 3 لأنه الأكثر تكرارًا.

1.2 - مقاييس التشتت(Measures of Dispersion)

يستخدم علماء الإحصاء عدة مقاييس لتحديد درجة انحراف البيانات عن القيمة الوسطية أي لقياس مدى انتشار البيانات حول القيمة المركزية، ويطلقون عليها اسم مقاييس التشتت، ومن أكثرها شيوعاً ما يلي: المدى، الانحراف المعياري.، التباين.

المدى(Range)

الفرق بين أكبر وأصغر قيمة في البيانات.

المدى = أكبر مشاهدة – أصغر مشاهدة.

اما في التوزيعات التكرارية يكون:

المدى = الحد الفعلي الأعلى للفئة العليا - الحد الفعلي الأدنى للفئة الدنيا

مثال: إذا كانت البيانات (5، 10، 15، 20، 25)، فإن:

المدى=25−5=20

التباين **(Variance- S^2)** : يُقيس مدى تشتت القيم حول المتوسط الحسابي.

المعادلة لعينة:

$$\hat{\sigma}^2 = \frac{1}{n-1} \cdot \sum_{i=1}^{n} (x_i - \bar{X})^2$$

ملاحظة:

التباين (كمفهوم في الإحصاء الوصفي) يساوي ببساطة المتوسط الحسابي لمربعات الفروق بين القيم والمتوسط الحسابي المُلاحظ.

أما التباين غير المتحيز (كمفهوم في الإحصاء الرياضي، والذي يعني أن القيمة التجريبية تساوي في المتوسط القيمة النظرية) فهو $\frac{n}{n-1}$ مرة قيمة التباين المُلاحظ، لذلك، يكون التباين غير المتحيز أكبر من التباين المُلاحظ.

مثال عملي:

إذا كانت درجات 4 طلاب : (10 ، 20، 30، 40).

المتوسط الحسابي : $\bar{x}=25$

التباين:

$$s^2 = \frac{(10-25)^2+(20-25)^2+(30-25)^2+(40-25)^2}{4-1} = 166.67$$

الانحراف المعياري (Standard Deviation – S) هو الجذر التربيعي للتباين ويُستخدم لقياس مدى انتشار البيانات حول المتوسط.

$$s = \sqrt{var} = \sqrt{s^2}$$

$$s = \sqrt{166.67} = 12.91$$

المعاملات الإحصائية (مثل معامل التشتت ومعامل التغير)

معامل التغير:(Coefficient of Variation - CV) : يُستخدم لمقارنة مدى التشتت بين مجموعات بيانات مختلفة.

$$C.V. = \frac{\sigma}{\bar{x}}$$

1.3 - التمثيل البياني للبيانات(Data Visualization)

التمثيل البياني يساعد على فهم طبيعة التوزيع واتجاه البيانات، ومن أهم الطرق:

المخطط التكراري:(Histogram) : يُستخدم لعرض توزيع التكرارات للبيانات المستمرة.

المحور الأفقي يمثل الفئات، والمحور العمودي يمثل التكرار.

المخطط الصندوقي:(Box Plot) : يُستخدم للكشف عن القيم الشاذة وتمثيل مقاييس النزعة المركزية والتشتت.

يعرض الوسيط، الربيع الأول، الربيع الثالث، والمدى بينهما.

المخطط الخطي:(Line Chart) يُستخدم لعرض الاتجاهات الزمنية.

أهمية التحليل أحادي المتغير

1. فهم طبيعة البيانات قبل الانتقال إلى التحليل الثنائي أو متعدد المتغيرات.
2. الكشف عن القيم الشاذة التي قد تؤثر على التحليل الإحصائي.
3. تحديد التوزيع الاحتمالي للمتغير (طبيعي، منحرف، متماثل).
4. إجراء التحليلات الوصفية لتقديم رؤى واضحة حول البيانات.

خلاصة:

التحليل أحادي المتغير يركز على دراسة متغير واحد باستخدام مقاييس النزعة المركزية، مقاييس التشتت، والتمثيل البياني.

يُعتبر خطوة أساسية لفهم البيانات قبل الانتقال إلى التحليل الثنائي أو متعدد المتغيرات.

تحليل ثنائي المتغير

التحليل ثنائي المتغير هو تحليل إحصائي يُستخدم لدراسة العلاقة بين متغيرين اثنين في مجموعة بيانات. يهدف هذا التحليل إلى فهم طبيعة العلاقة بين المتغيرات، سواء كانت علاقة ارتباطية (Correlation)أو علاقة سببية (Causation) .

وينقسم تحليل ثنائي المتغير الى:

أولا: تحليل العلاقة بين متغيرين كميين (عددين): يُستخدم عندما يكون كلا المتغيرين كميين ((Numerical، أي يمكن قياسهما بعدد معين.، ونستخدم لقياس العلاقة بينهما :

1. معامل الارتباط لبيرسون – Pearson's Correlation Coefficient)
2. تحليل الانحدار الخطي البسيط(Simple Linear Regression) : يُستخدم لإنشاء نموذج رياضي يصف العلاقة بين متغير مستقل X ومتغير تابع Y

ثانيا: تحليل العلاقة بين متغير كمي ومتغير نوعي: يُستخدم عندما يكون أحد المتغيرين كميًا (رقميًا) والآخر نوعيًا (تصنيفيًا أو فئويًا - .(Categorical، ونستعمل عادة:

1. **اختبار الفرق بين المجموعات(T-test & ANOVA) : مثل :**

❖ اختبار T المستقل :(Independent T-test) يُستخدم لمقارنة متوسط متغير كمي بين مجموعتين.

مثال: مقارنة متوسط الرواتب بين الذكور والإناث.

❖ تحليل التباين الأحادي :(ANOVA - Analysis of Variance) يُستخدم عندما يكون لدينا أكثر من مجموعتين.

مثال: مقارنة متوسط درجات الطلاب بين ثلاث مدارس مختلفة.

ثالثا: تحليل العلاقة بين متغيرين نوعيين : يُستخدم عندما يكون كلا المتغيرين تصنيفيين (.Categorical)

ومن بين الاختبارات المستخدمة :

اختبار مربع كاي χ^2- Chi-Square Test) : يُستخدم لاختبار العلاقة بين متغيرين نوعيين.

أهمية التحليل ثنائي المتغير

2. يُساعد في اكتشاف العلاقات بين المتغيرات لتحديد ما إذا كان هناك ارتباط قوي بينها.
3. يُستخدم في بناء النماذج التنبؤية، مثل الانحدار الخطي البسيط.
4. يُساعد في اختبار الفرضيات الإحصائية، مثل اختبار T واختبار مربع كاي.
5. يُستخدم في اتخاذ القرارات بناءً على تحليل البيانات، مثل تأثير التسويق على المبيعات.

خلاصة:

التحليل ثنائي المتغير يدرس العلاقة بين متغيرين فقط.
إذا كان كلاهما كميًا ← نستخدم معامل الارتباط أو الانحدار الخطي.
إذا كان أحدهما كميًا والآخر نوعيًا ← نستخدم T-test أو.ANOVA
إذا كان كلاهما نوعيًا ← نستخدم اختبار مربع كاي

التحليل الإحصائي متعدد المتغيرات (Multivariate Analysis)

التحليل متعدد المتغيرات هو فرع من الإحصاء يهدف إلى دراسة العلاقة بين أكثر من متغيرين في نفس الوقت. يُستخدم هذا التحليل عندما يكون لدينا مجموعة بيانات تحتوي على عدة متغيرات مرتبطة ببعضها، ويريد الباحث فهم العلاقات بينها واستخلاص استنتاجات ذات مغزى.

أنواع التحليل متعدد المتغيرات:

تحليل الانحدار المتعدد (Multiple Regression Analysis): يُستخدم لدراسة تأثير عدة متغيرات مستقلة على متغير تابع واحد.

تحليل الارتباط المتعدد Multiple Correlation Analysis: يُستخدم لقياس مدى ارتباط متغير تابع واحد بمجموعة من المتغيرات المستقلة.

التحليل العاملي Factor Analysis : هو تقنية إحصائية تُستخدم لاكتشاف العوامل الخفية (Latent Factors)التي تؤثر على مجموعة من المتغيرات الملاحظة. يساعد في تقليل عدد المتغيرات وتحديد البُنى الأساسية التي تفسر العلاقات بينها.

يُستخدم التحليل العاملي في تحليل الاستبيانات، علم النفس، التسويق، والاقتصاد لاكتشاف الأبعاد الرئيسية للبيانات، و يشمل التقنيات التالية:

1. تحليل المكونات الرئيسية (Principal Component Analysis - PCA)
2. التحليل التمييزي (Discriminant Analysis - DA)
3. تحليل التجميع (Cluster Analysis)

التمثيل البياني في التحليل متعدد المتغيرات

1. **المصفوفة الارتباطية: (Correlation Matrix)** تُستخدم لمعرفة العلاقات بين جميع المتغيرات.
2. **المخططات ثلاثية الأبعاد: (3D Scatter Plots)** تُستخدم لتمثيل العلاقة بين ثلاثة متغيرات.
3. **المخططات العنقودية (:(Cluster Plots** تُستخدم لعرض التجمعات المختلفة في تحليل التجميع.

أهمية التحليل متعدد المتغيرات

1. تحليل العلاقات المعقدة بين عدة متغيرات في وقت واحد.
2. تحسين دقة التنبؤات والنماذج الإحصائية.
3. استخلاص معلومات ذات معنى من البيانات الكبيرة.
4. تقليل حجم البيانات مع الاحتفاظ بأهم المعلومات من خلال تقنيات مثل.PCA

خلاصة:

التحليل متعدد المتغيرات يُستخدم عندما يكون هناك أكثر من متغيرين.
يشمل تقنيات مثل الانحدار المتعدد، تحليل المكونات الرئيسية، وتحليل العوامل.

مهم جدًا في الاقتصاد، التسويق، العلوم الاجتماعية، وتحليل البيانات الكبيرة.

الفصل الثاني : مراجعة حول الحساب المصفوفي

أولاً: ما هي المصفوفة؟

المصفوفة (Matrix) هي ترتيب مستطيل (أو مربع) من الأعداد (أو الرموز أو التعبيرات)، تُكتب على شكل صفوف وأعمدة.

$$\Delta = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$$

وتسمى الكميات a_{ik} المكونة للمصفوفة بعناصر المصفوفة حيث i عدد الصفوف و k عدد الأعمدة

وتسمى المصفوفة التى تحتوى على صفوف عددها m وأعمدة عددها n بمصفوفة ذات رتبة ($m\times n$) وإذا كانت m=n فإن المصفوفة تكون مربعة ولا توجد للمصفوفات أى قيمة جبرية بعكس المحددات.

مثال

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$$

هذه مصفوفة من الرتبة 2×3 (صفان و3 أعمدة).

أنواع المصفوفات الشائعة:

النوع	الوصف	مثال
مصفوفة مربعة	عدد الصفوف = عدد الأعمدة	$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$
مصفوفة صفية	صف واحد فقط	$[1\ 2\ 3]$
مصفوفة عمودية	عمود واحد فقط	$\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$
مصفوفة صفرية	كل عناصرها = 0	$\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$
مصفوفة وحدة I_n	مربعة، عناصر القطر الرئيسي = 1، والباقي = 0	$I_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$
مصفوفة قطرية	عناصر خارج القطر الرئيسي = 0	$\begin{bmatrix} 2 & 0 \\ 0 & 5 \end{bmatrix}$

مصفوفة مثلثية عليا/سفلى	العناصر تحت/فوق القطر = 0	(عليا) $\begin{bmatrix} 1 & 2 \\ 0 & 3 \end{bmatrix}$
مصفوفة متماثلة	$A = A^T$	$\begin{bmatrix} 1 & 2 \\ 2 & 3 \end{bmatrix}$
مصفوفة معكوسة	A^{-1} بحيث $AA^{-1} = I$ (توجد فقط للمصفوفات غير المفردة)	

أثر المصفوفة: Trace of a Matrix

أثر المصفوفة (يُرمز له بـــ tr(A) أو Tr(A)) هو مجموع عناصر القطر الرئيسي لمصفوفة مربعة A من الرتبة n×n :

$$\mathrm{tr}(A) = a_{11} + a_{22} + \cdots + a_{nn} = \sum_{i=1}^{n} a_{ii}$$

ملاحظة: يُعرّف الأثر فقط للمصفوفات المربعة.

أمثلة

$$A = \begin{bmatrix} 2 & 5 \\ -1 & 3 \end{bmatrix} \quad \Rightarrow \quad \mathrm{tr}(A) = 2 + 3 = 5$$

$$B = \begin{bmatrix} 4 & 1 & 0 \\ 2 & -1 & 3 \\ 0 & 5 & 2 \end{bmatrix} \quad \Rightarrow \quad \mathrm{tr}(B) = 4 + (-1) + 2 = 5$$

ملاحظات مهمة:

لا يُعرّف أثر المصفوفة غير المربعة (مثل 2×3) — لأن القطر ليس مكتملًا.

المصفوفة الصفرية. tr(0)=0 :

مصفوفة الوحدة I_n : $n = \mathrm{tr}(I_n)$

خصائص مهمة لأثر المصفوفة:

الخاصية	التعبير الرياضي	ملاحظات
الخطية	$\text{tr}(A+B) = \text{tr}(A) + \text{tr}(B)$	جمع مصفوفتين مربعتين من نفس الرتبة
	$\text{tr}(cA) = c \cdot \text{tr}(A)$	c عدد حقيقي أو مركب
الدوران تحت الضرب	$\text{tr}(AB) = \text{tr}(BA)$	حتى لو كانت $AB \neq BA$، فإن **أثرهما متساوٍ**!
	$\text{tr}(ABC) = \text{tr}(CAB) = \text{tr}(BCA)$	يمكن "تدوير" العوامل (لكن لا يمكن تبديل ترتيب عشوائي)
ثبات تحت التشابه	$\text{tr}(P^{-1}AP) = \text{tr}(A)$	الأثر لا يتغير بتغيير الأساس (مهم في القيم الذاتية)
علاقة بالقيم الذاتية	$\text{tr}(A) = \lambda_1 + \lambda_2 + \cdots + \lambda_n$	مجموع القيم الذاتية (مع العدّ حسب التعدد الجبري)
علاقة بالمحدّد	لا علاقة مباشرة، لكن: $\det(e^A) = e^{\text{tr}(A)}$	مفيدة في التحليل والفيزياء

العمليات الأساسية على المصفوفات: Matrices Algebra

جمع / طرح المصفوفات: Matrices addition and subtraction

يمكن إجراء عمليات الجمع والطرح للمصفوفات التي لها نفس الرتبة أي لها نفس العدد من الصفوف ونفس العدد من الأعمدة. فإذا كانت A , B مصفوفتان من نفس الرتبة فإن مجموع هاتين المصفوفتين يعرف بأنه يساوى المصفوفة C التي لها نفس الرتبة وكل عنصر من عناصرها يساوى مجموع العنصرين المتناظرين في A , B.

- يُطبق فقط إذا كانت المصفوفات من نفس الرتبة.
- يتم جمع/طرح العناصر المتناظرة.

مثال (1-3): إذا كانت:

$$A = \begin{pmatrix} 1 & -2 & 1 \\ 3 & 0 & 5 \end{pmatrix} \quad , \quad B = \begin{pmatrix} 3 & 6 & 5 \\ 2 & 2 & -1 \end{pmatrix}$$

$$\therefore C = A + B = \begin{pmatrix} 1+3 & -2+6 & 1+5 \\ 3+2 & 0+2 & 5-1 \end{pmatrix} = \begin{pmatrix} 4 & 4 & 6 \\ 5 & 2 & 4 \end{pmatrix}$$

$$D = A - B = \begin{pmatrix} 1-3 & -2-6 & 1-5 \\ 3-2 & 0-2 & 5+1 \end{pmatrix} = \begin{pmatrix} -2 & -8 & -4 \\ 1 & -2 & 6 \end{pmatrix}$$

ومن التعريف يمكن إثبات أن مجموع المصفوفات له الخصائص التالية:

$$A + B = B + A$$
$$A + (B + C) = (A + B) + C = A + B + C$$

وإذا كانت المصفوفة B هي حاصل جمع عدد من m من المصفوفات A فإن B = mA وكل عنصر من عناصر المصفوفة B يساوى m مضروبة في العنصر المقابل في المصفوفة A.

مثال

$$\begin{pmatrix} 2 & 5 \\ 3 & 4 \end{pmatrix} + \begin{pmatrix} 2 & 5 \\ 3 & 4 \end{pmatrix} = 2\begin{pmatrix} 2 & 5 \\ 3 & 4 \end{pmatrix} = \begin{pmatrix} 4 & 10 \\ 6 & 8 \end{pmatrix}$$

ومن هذا المثال يتضح الآتي:

1- تتساوى المصفوفتان A , B إذا تساوت رتبتاهما وتكون جميع عناصر كل منهما المتناظرة متساوية.

2- حاصل ضرب مصفوفة A في عدد m حقيقي أو تخيلي (مقدار قياسي أو مقدار ثابت) هو مصفوفة B عناصرها عبارة عن حاصل ضرب كل عنصر من عناصر A فى m. فمثلاً إذا كانت:

$$A = \begin{pmatrix} 1 & 3 & 2 \\ 2 & -1 & 0 \end{pmatrix} \qquad \therefore mA = \begin{pmatrix} m & 3m & 2m \\ 2m & -m & 0 \end{pmatrix}$$

ضرب المصفوفات:

شرط: عدد أعمدة المصفوفة الأولى = عدد صفوف المصفوفة الثانية.

الناتج: مصفوفة رتبتها (صفوف الأولى × أعمدة الثانية).

ليس تبديليًا عمومًا: $AB \neq BA$ في الغالب.

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}, \quad B = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \Rightarrow AB = \begin{bmatrix} 2 & 1 \\ 4 & 3 \end{bmatrix}, \quad BA = \begin{bmatrix} 3 & 4 \\ 1 & 2 \end{bmatrix}$$

مثال:

$$A=\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{pmatrix} \quad , \quad B=\begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix}$$

حاصل الضرب C = AB هو مصفوفة رتبتها (3 x 2) تعرف كالآتي:

$$C=\begin{pmatrix} a_{11}b_{11}+a_{12}b_{21} & a_{11}b_{12}+a_{12}b_{22} \\ a_{21}b_{11}+a_{22}b_{21} & a_{21}b_{12}+a_{22}b_{22} \\ a_{31}b_{11}+a_{32}b_{21} & a_{31}b_{12}+a_{32}b_{22} \end{pmatrix}$$

مثال: إوجد حاصل ضرب المصفوفتين:

$$A=\begin{pmatrix} 2 & 3 & -1 \\ 4 & 1 & 2 \end{pmatrix} \; , \; B=\begin{pmatrix} 2 & 1 \\ 4 & -2 \\ 5 & -3 \end{pmatrix}$$

الحـل

$$\begin{aligned} C &= A\times B \\ &= \begin{pmatrix} 2\times 2+3\times 4+(-1\times 5) & 2\times 1+3\times -2+(-1\times -3) \\ 4\times 2+1\times 4+2\times 5 & 4\times 1+1\times -2+2\times -3 \end{pmatrix} \\ &= \begin{pmatrix} 11 & -1 \\ 22 & -4 \end{pmatrix} \end{aligned}$$

وضرب المصفوفات له الخصائص التالية:

$$(1)\; A(B+C)=AB+AC$$
$$(2)\; (A+B)C=AC+BC$$
$$(3)\; A(BC)=AB(C)$$

. المنقول. (Le transpose)

يُرمز له بـ A^T ، ويُحصل عليه بتبديل الصفوف بالأعمدة.

$$A=\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \Rightarrow A^T=\begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix}$$

5. المحدّد (Determinant)

- يُعرف فقط للمصفوفات المربعة.
- يُرمز له بـ det(A) أو. |A|
- إذا $\det(A) \neq 0$ ⇦ المصفوفة قابلة للعكس (غير مفردة).
- لمصفوفة 2×2 :

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc,$$

لمصفوفة 3×3 :

$$\begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} = aei + bfg + cdh - ceg - bdi - afh.$$

أمثلة:

$$\begin{vmatrix} 1 & 2 \\ 1 & 3 \end{vmatrix} = 1 \cdot 3 - 2 \cdot 1 = 1$$

محدد مصفوفة n*n : يُحدد محدد مصفوفة ذات بُعد ما باستعمال صيغة لايبنتس أو صيغة لابلاس.

صيغة لايبنتس من أجل حساب محدد مصفوفة A بعدها n × n تأتي فيما يلي:

$$\det(A) = \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) \prod_{i=1}^{n} a_{i,\sigma_i}$$

الخواص محدد المصفوفة:

إذا كان A , B مصفوفتان مربعتان وقابلتان للضرب فإن:

1) $|AB| = |A|.|B|$

2) $|A_T| = |A|$

3) $|A^{-1}| = \dfrac{1}{|A|}$

حيث A^{-1} هو مقلوب أو معكوس المصفوفة كما سيعرف فيما بعد.

4) $|cA| = c^n.|A|$

حيث c مقدار ثابت، cA هو المصفوفة الناتجة من ضرب كل عنصر من عناصر المصفوفة A في المقدار الثابت c كما يلي:

$$cA = \begin{pmatrix} ca_{11} & ca_{12} & \cdots\cdots & \cdots\cdots & ca_{1n} \\ ca_{21} & ca_{22} & \cdots\cdots & \cdots\cdots & ca_{2n} \\ \cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots\cdots \\ \cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots\cdots \\ ca_{n1} & ca_{n2} & \cdots\cdots & \cdots\cdots & ca_{nn} \end{pmatrix}$$

$$\therefore |cA| = c \times c \times c \ldots\ldots \times c.|A| = c^n.|A|$$

5. المصفوفة العكسية A^{-1} :

- توجد فقط إذا كانت A مربعة وغير مفردة ($0 \neq \det(A)$).
- تحقق: $I = A^{1-}A = {}^{1-}AA$

مثال : لتكن المصفوفة التالية ذات البُعد الثاني: $\mathbf{A} = \begin{pmatrix} -1 & \frac{3}{2} \\ 1 & -1 \end{pmatrix}.$

هذه المصفوفة قابلة للعكس لكون محددها مختلفا عن الصفر

$\det \mathbf{A} = (-1 * -1) - (3/2 * 1) = -1/2$.

عكس مصفوفة باستعمال المصفوفة المر افقة والمحددة

يمكن إيجاد معكوس المصفوفة من القانون التالي:

$$A^{-1} = \frac{adj(A)}{|A|} \qquad A^{-1} = \frac{1}{\det A}\,{}^{t}\mathrm{com}\,A.$$

$$\therefore adj.A = B_T = \begin{pmatrix} A_{11} & A_{21} & \cdots & A_{n1} \\ A_{12} & A_{22} & \cdots & A_{n2} \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ A_{1n} & A_{2n} & \cdots & A_{nn} \end{pmatrix}$$

مصفوفة 2×2

$$\operatorname{com}\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

مصفوفة 3×3

$$\operatorname{com}\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} = \begin{pmatrix} +\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} & -\begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} & +\begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix} \\ -\begin{vmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{vmatrix} & +\begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{vmatrix} & -\begin{vmatrix} a_{11} & a_{12} \\ a_{31} & a_{32} \end{vmatrix} \\ +\begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix} & -\begin{vmatrix} a_{11} & a_{13} \\ a_{21} & a_{23} \end{vmatrix} & +\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} \end{pmatrix}$$

باستخدام حذف غاوسي: Gaussian elimination

هو خوارزمية مفيدة لحل منظومات من المعادلات الخطية وإيجاد رتبة مصفوفة وحساب معكوس مصفوفة مربعة انعكاسية، تم إعطاء هذا الاسم تقديرا للرياضياتي الألماني كارل فريدريك غاوس. يتم تطبيق عمليات الصف الأساسية لتخفيض المصفوفة على صورة مصفوفة مثلثية. يمكن تعميم هذه الخوارزمية باستخدام حذف غاوس جوردان، لتخفيض المصفوفة إلى صورة مصفوفة مثلثية مخفضة ومع ذلك فإن استعمال الحذف الغاوسي بمفرده كاف لأي تطبيق.

مثال:

$$\left[\begin{array}{ccc|c} 1 & 3 & 1 & 9 \\ 1 & 1 & -1 & 1 \\ 3 & 11 & 5 & 35 \end{array}\right] \to \left[\begin{array}{ccc|c} 1 & 3 & 1 & 9 \\ 0 & -2 & -2 & -8 \\ 0 & 2 & 2 & 8 \end{array}\right] \to \left[\begin{array}{ccc|c} 1 & 3 & 1 & 9 \\ 0 & -2 & -2 & -8 \\ 0 & 0 & 0 & 0 \end{array}\right] \to \left[\begin{array}{ccc|c} 1 & 0 & -2 & -3 \\ 0 & 1 & 1 & 4 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

مثال: إذا كانت:

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 5 \\ 1 & 5 & 12 \end{pmatrix}$$

إوجد مصفوفة العوامل المرافقة ثم إوجد المصفوفة البديلة.

الحـل

العامل المرافق للعنصر a11 هو المحددة:

$$A_{11} = \begin{pmatrix} 3 & 5 \\ 5 & 12 \end{pmatrix} = 11$$

والعامل المرافق للعنصر a12 هو المحددة:

$$A_{12} = -\begin{pmatrix} 1 & 5 \\ 1 & 12 \end{pmatrix} = -7$$

وهكذا....

وبذلك نحصل على مصفوفة العوامل المرافقة B ونجد أن:

$$B = \begin{pmatrix} 11 & -7 & 2 \\ -9 & 9 & -3 \\ 1 & -2 & 1 \end{pmatrix}$$

$$\therefore adj.A = B_T = \begin{pmatrix} 11 & -9 & 1 \\ -7 & 9 & -2 \\ 2 & -3 & 1 \end{pmatrix}$$

مثال : إوجد قيمة A^{-1} إذا كانت: $A = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 5 \\ 1 & 5 & 12 \end{pmatrix}$

الحـل

حسب المثال السابق فإن:

$$adj\ A=\begin{pmatrix} 11 & -9 & 1 \\ -7 & 9 & -2 \\ 2 & -3 & 1 \end{pmatrix}$$

$$|A| = 11-2\times7+3\times2=3\neq0$$

$$\therefore A^{-1}=\frac{adj\ A}{|A|}=\begin{pmatrix} 11/3 & -9/3 & 1/3 \\ -7/3 & 9/3 & -2/3 \\ 2/3 & -3/3 & 1/3 \end{pmatrix}=\begin{pmatrix} 11/3 & -3 & 1/3 \\ -7/3 & 3 & -2/3 \\ 2/3 & -1 & 1/3 \end{pmatrix}$$

قيم ذاتية ومتجهات ذاتية

عند تطبيق تحويل خطي (ممثل بمصفوفة A) على متجه ما v ، فإن المتجه عادةً يتغير في الطول والاتجاه.

لكن في بعض الحالات الخاصة، قد يبقى الاتجاه ثابتًا ، أي أن المتجه لا يدور، بل يتمدد أو ينكمش فقط (أو ينعكس)،

- **المتجه الذاتي** هو متجه لا يتغير اتجاهه تحت التحويل الخطي. A
- **القيمة الذاتية** هي عامل التمدد/الانكماش (أو الانعكاس) الذي يُطبَّق على هذا المتجه.

و تكتب صيغته الرياضية كالتالي:

$$A\mathbf{v}=\lambda\mathbf{v}$$

يسمى العدد λ قيمة ذاتية Eigenvalue للمصفوفة A تقابل المتجه الذاتي Eigenvector x.

كيف نحسب القيم والمتجهات الذاتية؟

الخطوة 1: حل المعادلة المميزة(Characteristic Equation)

$$A\mathbf{v}=\lambda\mathbf{v} \;\Rightarrow\; A\mathbf{v}-\lambda\mathbf{v}=\mathbf{0} \;\Rightarrow\; (A-\lambda I)\mathbf{v}=\mathbf{0}$$

لكي يوجد حل غير صفري ($v\neq0$) يجب أن تكون المصفوفة $(A-\lambda I)$ مفردة (غير قابلة للعكس)، أي :

$$\det(A-\lambda I)=0$$

هذه تُسمى المعادلة المميزة، وهي معادلة كثيرة حدود (من الدرجة n إذا كانت A من رتبة n) ، جذور هذه المعادلة = القيم الذاتية.

الخطوة 2: إيجاد المتجهات الذاتية

لكل قيمة ذاتية λ_i ، نعوّض في:

$$(A - \lambda_i I)\mathbf{v} = \mathbf{0}$$

ثم نحل النظام الخطي لإيجاد فضاء الحل(Null Space) ، وكل متجه غير صفري في هذا الفضاء هو متجه ذاتي مرتبط بـ λ

المتجهات الذاتية ليست وحيدة: أي مضاعف عددي غير صفري لمتجه ذاتي هو أيضًا متجه ذاتي (لأن الاتجاه هو المهم).

مثال:

$$A = \begin{bmatrix} 4 & 1 \\ 2 & 3 \end{bmatrix}$$

$$A - \lambda I = \begin{bmatrix} 4-\lambda & 1 \\ 2 & 3-\lambda \end{bmatrix}$$

$$\det(A - \lambda I) = (4-\lambda)(3-\lambda) - (1)(2) = \lambda^2 - 7\lambda + 10$$

$$\lambda^2 - 7\lambda + 10 = 0 \Rightarrow (\lambda - 5)(\lambda - 2) = 0 \quad \Rightarrow \quad \lambda_1 = 5, \quad \lambda_2 = 2$$

توجد المتجهات الذاتية:

$\lambda_1 = 5$

$$(A - 5I)\mathbf{v} = \begin{bmatrix} -1 & 1 \\ 2 & -2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

من الصف الأول : $-x + y = 0 \Rightarrow y = x$

إذن المتجهات الذاتية هي من الشكل: $\mathbf{v}_1 = t \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad t \neq 0$

$\lambda_2 = 2$

$$(A - 2I) = \begin{bmatrix} 2 & 1 \\ 2 & 1 \end{bmatrix} \Rightarrow 2x + y = 0 \Rightarrow y = -2x$$

$$\mathbf{v}_2 = s \begin{bmatrix} 1 \\ -2 \end{bmatrix}, \quad s \neq 0$$

لـ $\lambda = 5$: متجه ذاتي $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$

لـ $\lambda = 2$: متجه ذاتي $\begin{bmatrix} 1 \\ 2- \end{bmatrix}$

تقطير المصفوفة:

تقطير مصفوفة مربعة A يعني إيجاد مصفوفة قطرية D (أي جميع عناصرها خارج القطر الرئيسي تساوي صفرًا) ومصفوفة قابلة للعكس P بحيث:

$$\boxed{A = PDP^{-1}}$$

أو ما يعادلها

$$D = P^{-1}AP$$

نقول عندئذ: "المصفوفة A قابلة للتقطير" .(Diagonalizable)

متى تكون المصفوفة قابلة للتقطير؟

ليست كل المصفوفات قابلة للتقطير، تتمثل الشروط الأساسية في:

❖ **الشرط الضروري والكافي:**

المصفوفة A من الرتبة (n×n) قابلة للتقطير إذا وفقط إذا وُجدت n متجهات ذاتية مستقلة خطيًا .

❖ **ومن الشروط الكافية (أسهل تحققًا):**

1) إذا كانت لـA n قيم ذاتية مختلفة (أي جذور المعادلة المميزة متميزة)، فهي بالتأكيد قابلة للتقطير.

2) إذا كانت A متماثلة ($A=A^T$) ، فهي قابلة دائمًا للتقطير

3) وجود قيم ذاتية مكررة لا يعني بالضرورة أنها غير قابلة للتقطير، بل يعتمد على ما إذا كان التعدد الهندسي = التعدد الجبري لكل قيمة ذاتية.

خطوات تقطير مصفوفة

لتكن A مصفوفة مربعة . n×n

1 .أوجد القيم الذاتية: حل المعادلة المميزة:

$$\det(A - \lambda I) = 0$$

الحلول $\lambda_1, \lambda_2, \ldots, \lambda_k$. : (مع احتمال التكرار)

2. أوجد المتجهات الذاتية:

لكل قيمة ذاتية λ_i ، حل:

$$(A - \lambda_i I)\mathbf{v} = \mathbf{0}$$

واجمع جميع المتجهات الذاتية المستقلة خطيًا.

إذا حصلت على أقل من n متجهًا مستقلًا معناه ان A غير قابلة للتقطير.

3. شكّل المصفوفة P

ضع المتجهات الذاتية (المستقلة) كأعمدة في مصفوفة : P

$$P = \begin{bmatrix} | & | & & | \\ \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n \\ | & | & & | \end{bmatrix}$$

4. شكّل المصفوفة القطرية D

$$D = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}$$

5. تحقق من أن $A=PDP^{-1}$

مثال :

$$A = \begin{bmatrix} 4 & 1 \\ 2 & 3 \end{bmatrix}$$

القيم الذاتية: $\lambda_1 = 5$, $\lambda_2 = 2$

المتجهات الذاتية:

$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ (لـ $\lambda = 5$)،

$\mathbf{v}_2 = \begin{bmatrix} 1 \\ -2 \end{bmatrix}$ (لـ $\lambda = 2$)

$$P = \begin{bmatrix} 1 & 1 \\ 1 & -2 \end{bmatrix} \qquad D = \begin{bmatrix} 5 & 0 \\ 0 & 2 \end{bmatrix}$$

❖ **نحسب P^{-1}**

$$\det(P) = (1)(-2) - (1)(1) = -3$$

$$P^{-1} = \frac{1}{-3} \begin{bmatrix} -2 & -1 \\ -1 & 1 \end{bmatrix} = \begin{bmatrix} \frac{2}{3} & \frac{1}{3} \\ \frac{1}{3} & -\frac{1}{3} \end{bmatrix}$$

❖ **نتحقق من $A=PDP^{-1}$**

$$PD=\begin{bmatrix}1&1\\1&-2\end{bmatrix}\begin{bmatrix}5&0\\0&2\end{bmatrix}=\begin{bmatrix}5&2\\5&-4\end{bmatrix}$$

$$(PD)P^{-1}=\begin{bmatrix}5&2\\5&-4\end{bmatrix}\begin{bmatrix}\frac{2}{3}&\frac{1}{3}\\\frac{1}{3}&-\frac{1}{3}\end{bmatrix}=\begin{bmatrix}\frac{10+2}{3}&\frac{5-2}{3}\\\frac{10-4}{3}&\frac{5+4}{3}\end{bmatrix}=\begin{bmatrix}4&1\\2&3\end{bmatrix}=A$$

أهمية تقطير مصفوفة :

بمجرد أن تكتب $A=PDP^{-1}$ ، يصبح:

رفع المصفوفة لأس عال سهل جدا:

$$A^k=(PDP^{-1})^k=PD^kP^{-1}$$

وبما أن D قطرية، فإن:

$$D^k=\begin{bmatrix}\lambda_1^k&0&\cdots\\0&\lambda_2^k&\cdots\\\vdots&\vdots&\ddots\end{bmatrix}$$

مثال:

$$A^{100}=P\begin{bmatrix}5^{100}&0\\0&2^{100}\end{bmatrix}P^{-1}$$

(بدون حساب 100 ضرب مصفوفي!)

تمارين :

التمرين 1: لتكن:

$$A=\begin{bmatrix}2&-1\\0&3\end{bmatrix},\quad B=\begin{bmatrix}1&4\\-2&1\end{bmatrix},\quad C=\begin{bmatrix}0&2&-1\end{bmatrix}$$

احسب (إن أمكن):

1. $A+B$
2. $A-B$
3. $A+C$
4. $2A-3B$

التمرين 2: لتكن:

$$A=\begin{bmatrix}1 & 2\\ -1 & 0\end{bmatrix},\quad B=\begin{bmatrix}3 & -1\\ 2 & 4\end{bmatrix},\quad \mathbf{v}=\begin{bmatrix}2\\ -1\end{bmatrix}$$

احسب:

1. AB
2. BA
3. $A\mathbf{v}$
4. هل $AB=BA$? فسّر.

التمرين 3: احسب det(M) لكل مصفوفة:

1. $M_1=\begin{bmatrix}5 & 2\\ 1 & 3\end{bmatrix}$
2. $M_2=\begin{bmatrix}1 & 0 & 2\\ -1 & 3 & 1\\ 0 & 2 & -1\end{bmatrix}$
3. $M_3=\begin{bmatrix}a & b\\ c & d\end{bmatrix}$ (بصيغة عامة)
4. إذا كانت $\det(A)=4$ ، فما هو $\det(2A)$ إذا كانت A من رتبة 3×3 ؟

التمرين 4

1. أوجد معكوس المصفوفة:

$$A=\begin{bmatrix}4 & 7\\ 1 & 2\end{bmatrix}$$

2. هل المصفوفة التالية قابلة للعكس؟ ولماذا؟

$$B = \begin{bmatrix} 2 & 4 \\ 1 & 2 \end{bmatrix}$$

3. تحقق من أن : $AA^{-1} = I$.

التمرين 5:

$$A = \begin{bmatrix} 3 & 1 \\ 0 & 2 \end{bmatrix}$$

1. أوجد القيم الذاتية لـ A.
2. أوجد المتجهات الذاتية.
3. هل A قابلة للتقطير؟ علل.
4. إذا كانت قابلة، فاكتب. $A=PDP^{-1}$:

حل التمرين 1

$$A + B = \begin{bmatrix} 2+1 & -1+4 \\ 0+(-2) & 3+1 \end{bmatrix} = \begin{bmatrix} 3 & 3 \\ -2 & 4 \end{bmatrix}$$

$$A - B = \begin{bmatrix} 2-1 & -1-4 \\ 0-(-2) & 3-1 \end{bmatrix} = \begin{bmatrix} 1 & -5 \\ 2 & 2 \end{bmatrix}$$

2. غير معرف، لأن A رتبتها 2×2 و C رتبتها 1×3 ⟵ لا يمكن الجمع.

$$2A = \begin{bmatrix} 4 & -2 \\ 0 & 6 \end{bmatrix}, \quad 3B = \begin{bmatrix} 3 & 12 \\ -6 & 3 \end{bmatrix}$$

$$2A - 3B = \begin{bmatrix} 4-3 & -2-12 \\ 0-(-6) & 6-3 \end{bmatrix} = \begin{bmatrix} 1 & -14 \\ 6 & 3 \end{bmatrix}$$

حل التمرين 2

$$AB = \begin{bmatrix} 1\cdot 3 + 2\cdot 2 & 1\cdot(-1) + 2\cdot 4 \\ -1\cdot 3 + 0\cdot 2 & -1\cdot(-1) + 0\cdot 4 \end{bmatrix} = \begin{bmatrix} 7 & 7 \\ -3 & 1 \end{bmatrix}$$

$$BA = \begin{bmatrix} 3\cdot 1 + (-1)\cdot(-1) & 3\cdot 2 + (-1)\cdot 0 \\ 2\cdot 1 + 4\cdot(-1) & 2\cdot 2 + 4\cdot 0 \end{bmatrix} = \begin{bmatrix} 4 & 6 \\ -2 & 4 \end{bmatrix}$$

$$A\mathbf{v} = \begin{bmatrix} 1 & 2 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} 2 \\ -1 \end{bmatrix} = \begin{bmatrix} 1\cdot 2 + 2\cdot(-1) \\ -1\cdot 2 + 0\cdot(-1) \end{bmatrix} = \begin{bmatrix} 0 \\ -2 \end{bmatrix}$$

4. لا، لأن $AB = \begin{bmatrix} 7 & 7 \\ 1 & 3- \end{bmatrix}$ ، و$BA = \begin{bmatrix} 6 & 4 \\ 4 & 2- \end{bmatrix}$ ، وهما مختلفتان.

❖ ضرب المصفوفات ليس تبديليًا بشكل عام

حل التمرين 3

$$\det(M_1) = (5)(3) - (2)(1) = 15 - 2 = 13$$

نحسب بفك لابلاس حسب الصف الأول:

$$\det(M_2) = 1\cdot\det\begin{bmatrix} 3 & 1 \\ 2 & -1 \end{bmatrix} - 0\cdot(\cdots) + 2\cdot\det\begin{bmatrix} -1 & 3 \\ 0 & 2 \end{bmatrix}$$

$$= 1\cdot(3\cdot(-1) - 1\cdot 2) + 2\cdot((-1)\cdot 2 - 3\cdot 0) = 1\cdot(-5) + 2\cdot(-2) = -5 - 4 = -9$$

$$\det(M_3) = ad - bc$$

4. خاصية مهمة:

$$\det(kA) = k^n \det(A)$$ إذا كانت A من رتبة $n \times n$

هنا $n = 3$ ، لذا:

$$\det(2A) = 2^3 \cdot \det(A) = 8\cdot 4 = 32$$

حل التمرين 4

$$A = \begin{bmatrix} 4 & 7 \\ 1 & 2 \end{bmatrix}, \quad \det(A) = 4 \cdot 2 - 7 \cdot 1 = 8 - 7 = 1 \neq 0 \Rightarrow \text{قابلة للعكس}$$

للمصفوفة 2×2 :

$$A^{-1} = \frac{1}{\det(A)} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} = \frac{1}{1} \begin{bmatrix} 2 & -7 \\ -1 & 4 \end{bmatrix} = \begin{bmatrix} 2 & -7 \\ -1 & 4 \end{bmatrix}$$

$$\det(B) = 2 \cdot 2 - 4 \cdot 1 = 4 - 4 = 0 \Rightarrow \text{غير قابلة للعكس (مفردة)}$$

3. تحقق:

$$AA^{-1} = \begin{bmatrix} 4 & 7 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 2 & -7 \\ -1 & 4 \end{bmatrix} = \begin{bmatrix} 8-7 & -28+28 \\ 2-2 & -7+8 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I \quad \checkmark$$

حل التمرين 5

$$A = \begin{bmatrix} 3 & 1 \\ 0 & 2 \end{bmatrix}$$

1. **القيم الذاتية**:

$$\det(A - \lambda I) = \begin{vmatrix} 3-\lambda & 1 \\ 0 & 2-\lambda \end{vmatrix} = (3-\lambda)(2-\lambda) = 0 \Rightarrow \lambda_1 = 3, \ \lambda_2 = 2$$

قيم مختلفة → قابلة للتقطير.

2. **المتجهات الذاتية**:

- لـ $\lambda = 3$:

$$\begin{bmatrix} 1 \\ 0 \end{bmatrix} = {}_1\mathbf{v} \Rightarrow \text{حر } {}_1v \ ,0 = {}_2v \Rightarrow 0 = \mathbf{v} \begin{bmatrix} 1 & 0 \\ 1- & 0 \end{bmatrix} = 3I)\mathbf{v} - A)$$

- لـ $\lambda = 2$:

$$(\begin{bmatrix} 1- \\ 1 \end{bmatrix} \text{ أو}) \begin{bmatrix} 1 \\ 1- \end{bmatrix} = {}_2\mathbf{v} \Rightarrow {}_2v- = {}_1v \Rightarrow 0 = {}_2v + {}_1v \Rightarrow \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} = (2I - A)$$

3. وُجد متجهان مستقلان ⟶ قابلة للتقطير

.4

$$P = \begin{bmatrix} 1 & 1 \\ 0 & -1 \end{bmatrix}, \quad D = \begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix}$$

نحسب P^{-1} :

$\det(P) = (1)(-1) - (1)(0) = -1$

$$P^{-1} = \frac{1}{-1} \begin{bmatrix} -1 & -1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & -1 \end{bmatrix}$$

تحقق سريع:

$$PDP^{-1} = \begin{bmatrix} 1 & 1 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & -1 \end{bmatrix} = \begin{bmatrix} 3 & 2 \\ 0 & -2 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & -1 \end{bmatrix} = \begin{bmatrix} 3 & 1 \\ 0 & 2 \end{bmatrix} = A \quad \checkmark$$

الفصل الثالث : تحليل الارتباط و التباين المشترك

تعريف الارتباط:

يقصد بالارتباط وجود علاقة إحصائية بين متغيرين أو أكثر، بحيث إن تغير أحد المتغيرات يؤدي إلى تغير المتغير الآخر، إما في الاتجاه نفسه أو في الاتجاه المعاكس. فإذا زادت قيم أحد المتغيرين ورافقتها زيادة في قيم المتغير الآخر، سميت العلاقة علاقة طردية أو موجبة، أما إذا كانت زيادة أحد المتغيرين تقابلها نقصان في المتغير الآخر، فتُعرف العلاقة بالعلاقة العكسية أو السالبة.

ويستخدم الارتباط لدراسة درجة وقوة العلاقة بين الظاهرتين محل الدراسة، دون أن يعني ذلك بالضرورة وجود علاقة سببية مباشرة بينهما.

وتقاس قوة واتجاه العلاقة بين المتغيرات باستخدام مقياس كمي يُعرف بـ معامل الارتباط (Correlation Coefficient)، والذي يحدد:

1. اتجاه العلاقة (طردية أو عكسية).
2. درجة قوة العلاقة (ضعيفة، متوسطة، أو قوية).

أما عند دراسة تأثير متغير مستقل على متغير تابع بصورة أكثر تفصيلًا، فيتم استخدام تحليل الانحدار (Regression Analysis)، الذي يهدف إلى إيجاد علاقة رياضية أو جبرية بين المتغيرين، تُعرف بـ معادلة الانحدار. وفي حالة العلاقة الخطية البسيطة، تأخذ المعادلة الشكل التالي:

$$y=ax+b$$

حيث:

y : المتغير التابع.

x : المتغير المستقل.

a : ميل المستقيم ويمثل مقدار التغير في y الناتج عن تغير وحدة واحدة في.x

b :الجزء المقطوع من محور الصادات.

ويعد معامل الارتباط من أهم المقاييس الإحصائية المستخدمة في الدراسات الاقتصادية والاجتماعية والطبيعية، لما يوفره من وصف دقيق لطبيعة العلاقة بين المتغيرات.

1. معامل الارتباط الخطي Linear Correlation / بيرسون

يعرف معامل الارتباط الخطي البسيط بـ معامل ارتباط بيرسون (Pearson Correlation Coefficient)، ويُسمّى كذلك بالمعامل العزومي أو التتابعي، نسبةً إلى عالم الإحصاء البريطاني Karl Pearson الذي أسهم في تطوير هذا المقياس الإحصائي. ويستخدم هذا المعامل لقياس درجة وقوة واتجاه العلاقة الخطية بين متغيرين كميين، حيث تتراوح قيمته بين −1 و +1.

فإذا كانت قيمة المعامل موجبة دل ذلك على وجود علاقة طردية بين المتغيرين، أما إذا كانت سالبة فهذا يشير إلى وجود علاقة عكسية، في حين تعبّر القيمة القريبة من الصفر عن ضعف أو انعدام العلاقة الخطية بينهما:

1. إذا ازداد Y بازدياد X ⟵ ارتباط موجب($r > 0$)
2. إذا انخفض Y بازدياد X ⟵ ارتباط سالب($r < 0$)
3. إذا لم يتأثر Y بتغيرات X ⟵ انعدام الارتباط($r \approx 0$)

ولحساب معامل الارتباط لعيّنة مؤلفة من n زوجاً مرتباً : $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ يُستخدم أحد الصيغ الآتية:

$$r = \frac{\sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})}{(n-1)\, S_x\, S_y}$$

$$r = \frac{n \sum X_i Y_i - (\sum X_i)(\sum Y_i)}{\sqrt{n\sum X_i^2 - (\sum X_i)^2} \cdot \sqrt{n\sum Y_i^2 - (\sum Y_i)^2}}$$

$$r = \frac{\sum_{i=1}^{n} X_i Y_i - n\bar{X}\bar{Y}}{\sqrt{\sum X_i^2 - n\bar{X}^2} \cdot \sqrt{\sum Y_i^2 - n\bar{Y}^2}}$$

$$r = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{\sqrt{\sum (X - \bar{X})^2} \cdot \sqrt{\sum (Y - \bar{Y})^2}}$$

ويشترط لاستخدام معامل ارتباط بيرسون توفر مجموعة من الفرضيات الإحصائية الأساسية، أهمها:

1. أن يكون المتغيران محل الدراسة كميين (نسبيين أو فئويين) .
2. أن يتبع توزيع كل من المتغيرين التوزيع الطبيعي أو الاعتدالي .
3. أن تكون العينة مختارة بطريقة عشوائية .
4. استقلالية المشاهدات، أي أن قيمة كل فرد في العينة لا تعتمد على قيم الأفراد الآخرين .
5. وجود علاقة خطية تقريبية بين المتغيرين .

وفي حالة عدم تحقق شرط الاعتدالية، أو عندما تكون البيانات رتبية، يتم اللجوء إلى معاملات ارتباط لا معلمية(Nonparametric Correlation Measures) ، ومن أشهرها:

1. معامل ارتباط Charles Spearmanسبيرمان .(Spearman Rank Correlation)
2. معامل Maurice Kendallكندال تاو .(Kendall's Tau)

ويُعد معامل ارتباط بيرسون من أكثر المقاييس الإحصائية استخدامًا في الدراسات الاقتصادية والاجتماعية والإنسانية، لما يتميز به من بساطة في التفسير ودقة في قياس قوة العلاقة الخطية بين المتغيرات.

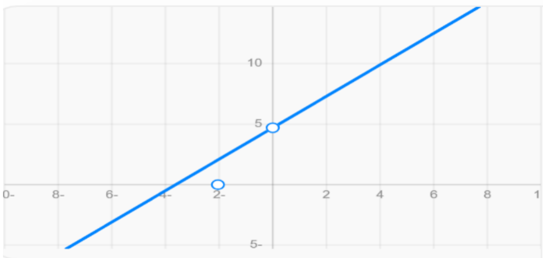
العوامل المؤثرة في معامل ارتباط بيرسون:

يتأثر معامل ارتباط بيرسون Karl Pearson بعدة عوامل إحصائية ومنهجية تؤثر في قوة العلاقة الخطية بين المتغيرين ودقة تقديرها، ومن أهم هذه العوامل ما يلي:

1. خطية العلاقة بين المتغيرين

يشترط في استخدام معامل ارتباط بيرسون أن تكون العلاقة بين المتغيرين علاقة خطية، أي يمكن تمثيلها تقريبًا بالمعادلة:

$$y = ax+b$$



ويُقصد بذلك أن تقع نقاط الأزواج المرتبة (x,y) على خط مستقيم أو قريبة جدًا منه، ويمكن التحقق من هذه الخاصية من خلال مخطط الانتشار (Scatter Plot). أما إذا كانت العلاقة غير خطية، فإن معامل بيرسون لا يكون مناسبًا، ويُفضل استعمال معاملات ارتباط أخرى أكثر ملاءمة لطبيعة البيانات.

2. مقدار التباين بين القيم

كلما وُجد تباين واضح ومنتظم في قيم المتغيرين، أصبحت العلاقة بينهما أكثر وضوحًا، مما يساهم في زيادة قوة معامل الارتباط. أما إذا كانت القيم متقاربة جدًا وضعيفة التشتت، فإن ذلك قد يؤدي إلى انخفاض قيمة معامل الارتباط.

3. حجم العينة الإحصائية

تزداد دقة وثبات معامل الارتباط بزيادة حجم العينة، إذ إن العينات الكبيرة تعطي تقديرات أكثر موثوقية وتمثيلًا للمجتمع الإحصائي، بينما قد تؤدي العينات الصغيرة إلى نتائج مضللة أو غير مستقرة.

4. شكل التوزيع وتماثله

يتأثر معامل الارتباط بشكل التوزيع الإحصائي للمتغيرين ودرجة تماثلهما:

- إذا كان شكل التوزيعين متماثلًا تمامًا وكانت العلاقة خطية قوية، فإن قيمة معامل الارتباط تقترب من1 : r=±1
- إذا كان التواء المتغيرين في الاتجاه نفسه، فإن العلاقة تكون طردية قوية وتقترب قيمة المعامل من : r=1
- أما إذا كان التواء أحد المتغيرين موجبًا والآخر سالبًا، فإن العلاقة تكون عكسية قوية وتقترب قيمة المعامل من : 1−=r

خصائص معامل ارتباط بيرسون

يتميز معامل ارتباط بيرسون بعدة خصائص إحصائية مهمة، من أبرزها:

1. أنه لا يعتمد على القيم المطلقة للمتغيرات، وإنما يعتمد على درجة تباعد القيم وتشتتها حول المتوسط الحسابي.
2. لا تتغير قيمة معامل الارتباط عند إجراء العمليات الحسابية الأساسية على القيم، مثل:
   - إضافة ثابت .
   - طرح ثابت .
   - الضرب في ثابت موجب .
   - القسمة على ثابت موجب .

وذلك لأن معامل الارتباط يقيس طبيعة العلاقة واتجاهها، وليس القيم العددية ذاتها للمتغيرات.

<u>مثال1:</u> لنفترض أننا نريد دراسة العلاقة بين عدد ساعات المراجعة ونتائج الطلبة في الاختبار، وكانت البيانات كما يلي:

علامة الاختبار(y)	عدد ساعات المراجعة(x)
50	2
55	3
65	4
70	5
80	6

من خلال ملاحظة البيانات نجد أنه كلما زاد عدد ساعات المراجعة ارتفعت علامات الطلبة، مما يدل مبدئيًا على وجود علاقة طردية بين المتغيرين.

حساب معامل الارتباط باستخدام العلاقة الز ائية:

$$Z = \frac{X-\bar{X}}{S}$$

1. حساب المتوسط الحسابي:

$$\bar{x} = \frac{2+3+4+5+6}{5} = 4 \qquad \bar{y} = \frac{50+55+65+70+80}{5} = 64$$

2. حساب الانحراف المعياري

$$S_x = \sqrt{\frac{(-2)^2+(-1)^2+0^2+1^2+2^2}{5}} = \sqrt{2} \approx 1.414 \qquad S_y = \sqrt{\frac{(-14)^2+(-9)^2+1^2+6^2+16^2}{5}} = \sqrt{114} \approx 10.677$$

x	y	Z_x	Z_y	Z_xZ_y
2	50	-1.414	-1.311	1.854
3	55	-0.707	-0.843	0.596
4	65	0	0.094	0
5	70	0.707	0.562	0.397
6	80	1.414	1.498	2.118

$\sum Z_xZ_y$ = **4.965**

حساب معامل الارتباط

$$r = \frac{\sum Z_x Z_y}{n}$$

$$r = \frac{4.965}{5} \approx 0.993$$

إذن معامل الارتباط يساوي تقريبًا : $r\approx0.99$ وهو يدل على وجود علاقة طردية قوية جدًا بين عدد ساعات المراجعة ونتائج الاختبار.

2. حساب معامل الارتباط من البيانات الأصلية

نستخدم القانون التالي:

$$r = \frac{n\sum xy - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}}$$

	x	y	x^2	y^2	xy
	2	50	4	2500	100
	3	55	9	3025	165
	4	65	16	4225	260
	5	70	25	4900	350
	6	80	36	6400	480
المجاميع	20	320	90	21050	1355

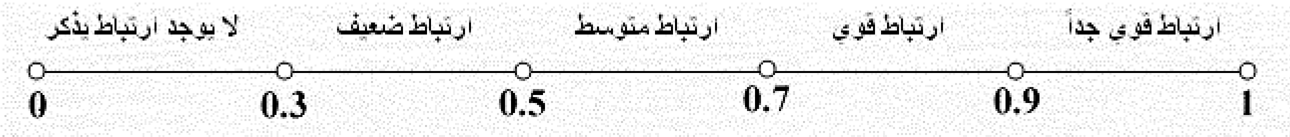
البسط: $5(1355) - (20)(320) = 375$

المقام: $\sqrt{[5(90) - 20^2]\,[5(21050) - 320^2]} = \sqrt{50 \times 28250}$

$\sqrt{1412500} \approx 375.83$

$$r = \frac{375}{375.83} \approx 0.998$$

قيمة معامل ارتباط Karl Pearson بيرسون قريبة جدًا من +1، وهذا يدل على وجود علاقة خطية طردية قوية جدا بين عدد ساعات المراجعة ونتائج الاختبار.



➢ **حساب معنوية معامل الارتباط**

بعد حساب معامل ارتباط Karl Pearson بيرسون، لا يكفي معرفة قيمة الارتباط فقط، بل يجب التحقق إحصائيًا مما إذا كانت هذه العلاقة معنوية (حقيقية) أو أنها ظهرت بالصدفة نتيجة العينة المستخدمة. ويتم ذلك من خلال اختبار معنوية معامل الارتباط.

من المثال السابق : **r=0.998** و حجم العينة: n=5

1) فرضيات الاختبار:

➢ **الفرضية العدمية :**

$H_0 : r = 0$

أي لا توجد علاقة ارتباط معنوية بين المتغيرين.

➢ **الفرضية البديلة :**

$H_1 : r \neq 0$

أي توجد علاقة ارتباط معنوية بين المتغيرين.

2) **حساب إحصائية الاختبار**

يتم اختبار معنوية معامل الارتباط باستخدام توزيع ستيودنت t لـ (William Sealy Gosset)، وفق العلاقة:

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

$$t = \frac{0.998\sqrt{3}}{\sqrt{0.004}} \approx \frac{1.728}{0.063} \approx 27.43$$

القيمة الجدولية

درجات الحرية:

df = n−2 = 3

وعند مستوى معنوية 5% واختبار ثنائي الطرف نجد من جدول t :

$$t_{0.05,3} = 3.182$$

بما أن:

$$3.182 < 27.43$$

فإننا نرفض الفرضية العدمية H_0 ،ونقبل الفرضية البديلة.

نستنتج أنه يوجد ارتباط خطي طردي معنوي إحصائيا بين عدد ساعات المراجعة ونتائج الاختبار، أي أن العلاقة بين المتغيرين ليست عشوائية بل علاقة حقيقية يمكن الاعتماد عليها في التفسير والتحليل الإحصائي.

تمرين:

يمثل الجدول التالي درجات 8 موظفين في اختبارين: المهارات التقنية (X) وأداء العمل (Y) :

الموظف	المهارات التقنية (X)	(Y) أداء العمل
1	65	70
2	72	68
3	80	85
4	58	60
5	90	92
6	75	78
7	68	65
8	85	88

المطلوب :

1. هل توجد علاقة بين المهارات التقنية (X) وأداء العمل (Y)
2. ماذا تستنتج عند مستوى معنوية 0.05؟

معامل ارتباط الرتب: (Rank Correlation Conefficient)

هذا المعامل يعرف بمعامل ارتباط سبيرمان (Spearman) أو معامل ارتباط الرتب (رتب القيم الأصلية وليس القيم) ولذا تختلف قيمته عن قيمة معامل بيرسون (للقيم الأصلية وليس لرتبها) وهو أقل دقة من معامل ارتباط بيرسون ويتعامل مع البيانات الرقمية وغير الرقمية للترتيب مثل جيد، جيد جدا، ... ويرمز له بالرمز r_s وهو ضمن الإحصاءات غير المعلمية ذات التوزيع الحر وقيمته موجبة أقل أو تساوي الواحد الصحيح وتحسب قيمته من الصيغة الرياضية علماً بأن:

$$r_s = 1 - \frac{6\sum d_i^2}{n(n^2 - 1)}$$

حيث d الفرق بين رتبه حسب المتغير الأول x ورتبه حسب المتغير الثاني) yالفرق بين رتب القيم لكل زوج من البيانات) وفي حالة التساوي يأخذ المتوسط الحسابي (فإذا كانت لقيمتين متساويتين الرتبتين 7 ، 8 فيأخذ متوسط 7 ، 8 وتصبح الرتب لكل منها 7.5 بدل عن 7 ، 8 (، nعدد الأزواج للقيم فإذا كان لدينا مجموعة من الأفراد وجرى ترتيبهم حسب صفتين لكل فرد من المجموعة x , y فإن. $d_i = x_i - y_i$

مثال:

تقدم عشرة طلاب لامتحان المرحلة الثانوية وكانت معدلات نتائجهم حسب الصف والمدرسة كالتالي والمطلوب حساب معامل سبيرمان للارتباط.

معدل الطالب في الصف (X)	73	80	70	66	89	71	65	88	92	74
مدل الطالب في المدرسة (Y)	69	78	66	70	92	64	55	90	88	72

الحل:

نكون جدول نبين فيه رتب كل من X (المعدل في الصف) و X (المعدل في المدرسة) والفرق d ومربع الفرق d^2 كالتالي:

X	Y	Rank X	Rank Y	d	d^2
73	69	6	7	– 1	1
80	78	4	4	0	0
70	66	8	8	0	0
66	70	9	6	3	9
89	92	2	1	1	1
71	64	7	9	– 2	4
65	55	10	10	0	0

88	90	3	2	1	1
92	88	1	3	– 2	4
74	72	5	5	0	0

بتطبيق القانون أعلاه:

$$r_s = 1 - \frac{6\sum d^2}{n(n^2-1)}$$

$$r_s = 1 - \frac{6 \times 20}{10(100-1)}$$

$$r_s = 1 - \frac{120}{990}$$

$$r_s = 1 - 0.121$$

$$r_s = 0.879$$

معامل الارتباط المتعدد:

هذا المعامل والذي يرمز له بالرمز R أيضاً يقيس قوة العلاقة بين أكثر من متغيرين وهي مغيرات عشوائية متصلة التوزيع (توزيع متعدد المتغيرات Multivariate distribution) وإن حساب قيمة R هو امتداد لقيمة معامل الارتباط البسيط (r) مع استبدال X ,Y بـ $X_1 - X_k$, Y ولنأخذ ثلاث متغيرات X_1، X_2، X_3 نحصل على الصيغ الآتية:

$$r_{12} = \frac{n\sum X_1X_2 - \sum X_1 \sum X_2}{\sqrt{n\sum X_1^2 - (\sum X_1)^2}\sqrt{n\sum X_2^2 - (\sum X_2)^2}}$$

$$r_{13} = \frac{n\sum X_1X_3 - \sum X_1 \sum X_3}{\sqrt{n\sum X_1^2 - (\sum X_1)^2}\sqrt{n\sum X_3^2 - (\sum X_3)^2}}$$

$$r_{23} = \frac{n\sum X_2X_3 - \sum X_2 \sum X_3}{\sqrt{n\sum X_2^2 - (\sum X_2)^2}\sqrt{n\sum X_3^2 - (\sum X_3)^2}}$$

يمكن حساب معامل الارتباط المتعدد من الصيغة:-

$$R_{123} = \sqrt{\frac{r_{12}^2 + r_{13}^2 - 2r_{12}r_{13}r_{23}}{1 - r_{23}^2}}$$

أو

$$R_{213} = \sqrt{\frac{r_{21}^2 + r_{23}^2 - 2r_{21}r_{13}r_{13}}{1 - r_{13}^2}}$$

ومعامل الارتباط المتعدد قيمته بين الصفروالواحد الصحيح وهو موجب دائماً

مثال:

أرادت مؤسسة للدعاية والإعلان معرفة العلاقة بين عدد المستجيبين لإعلاناتها y وحجم الإعلان المنشور في الصحيفة X_1 وعدد الصحف الموزعة X_2 التي تنشر الإعلان وحصلت المؤسسة على البيانات التالية:

عدد المستجيبين بالمئات (y_i) ، حجم الإعلان بالإنش (X_1) ، عدد الصحف الموزعة بالآلف (X_2)

X_2	X_2^2	X_1	X_1^2	y_i	y_i^2	y_iX_1	y_iX_2	X_1X_2
2	4	1	1	1	1	1	2	2
8	64	8	64	4	16	32	32	64
1	1	3	9	1	1	3	1	3
7	49	5	25	3	9	15	21	35
4	16	6	36	2	4	12	8	24
6	36	10	100	4	16	40	24	60
28	170	33	235	15	47	103	88	188

$$r_{y1}=\frac{n\sum yX_1-\sum y\sum X_1}{\sqrt{n\sum y^2-(\sum y)^2}\sqrt{n\sum X_1^2-(\sum X_1)^2}}=\frac{6\times103-15\times33}{\sqrt{6\times47-(15)^2}\sqrt{6\times235-(33)^2}}=0.909$$

$$r_{y2}=\frac{n\sum yX_2-\sum y\sum X_2}{\sqrt{n\sum y^2-(\sum y)^2}\sqrt{n\sum X_2^2-(\sum X_2)^2}}=\frac{6\times88-15\times28}{\sqrt{6\times47-(15)^2}\sqrt{6\times170-(28)^2}}=0.931$$

$$r_{12}=\frac{n\sum X_1X_2-\sum X_1\sum X_2}{\sqrt{n\sum X_1^2-(\sum X_1)^2}\sqrt{n\sum X_2^2-(\sum X_2)^2}}=\frac{6\times188-33\times28}{\sqrt{6\times235-(33)^2}\sqrt{6\times170-(28)^2}}=0.741$$

$$R_{y12}=\sqrt{\frac{r_{y1}^2+r_{y2}^2-2r_{y1}r_{y2}r_{12}}{1-r_{12}^2}}=\sqrt{\frac{0.826+0.867-2\times0.909\times0.931\times0.741}{1-0.549}}=0.973$$

اختبار فرضية العدم H_o : $P_{y.12...k}=0$ نحسب F من الصيغة:

$$F=\frac{R_{y.12..k}^2}{1-R_{y.12..k}^2}-\frac{n-k-1}{k}$$

حيث k عدد المتغيرات ونقارن قيمة F المحسوبة من الصيغة الرياضية السابقة مع قيمة F الجدولية فإن كانت أقل نقبل H_o

من الجدول نجد أن $F_{0.025,3,2}=16.04$ ومن الصيغة أعلاه نجد أن:

$$F = \frac{R^2_{y.12..k}}{1-R^2_{y.12..k}} - \frac{n-k-1}{k}$$

$$F = \frac{(0.973)^2}{1-(0.973)^2} - \frac{6-3-1}{3}$$

$$F = \frac{0.947}{1-0.947} - \frac{2}{3}$$

$$F = \frac{0.947}{0.053} - \frac{2}{3}$$

$$F = 17.2$$

وبما أن 17.2 < 16.04 فنرفض الفرضية الصفرية ونستدل على معنوية R

معامل الارتباط الجزئي:

هذا المعامل يقيس الارتباط بين أي زوج من المتغيرات عند ثبات المتغيرات الأخرى ويرمزله بالرمز $r_{12.345}$ فالارتباط هنا بين المتغيرين X_1 , X_2 وثبات الثلاثة الأخرى X_3 , X_4 , X_5 ويستخدم لتحديد العلاقة بين متغيرين محددين أو للتعرف على متغيرات يراد حذفها بسبب عدم تأثيرها على المتغير التابع ومربع معامل الارتباط الجزئي يعرف بمعامل التحديد الجزئي، ويجب ملاحظة أن r_{y1} رمز للعلاقة بين y , x_1 في حين $r_{y1.2}$ رمز للعلاقة بين y , x_1 مع ثبات x_2. يحسب معامل الارتباط الجزئي من الصيغة الرياضية الآتية (بين y , x_2 وثبات x_1):

$$r_{y2.1} = \frac{r_{y2} - r_{y1}\, r_{12}}{\sqrt{(1-r^2_{y1})(1-r^2_{12})}}$$

<u>مثال:</u>

أوجد معامل الارتباط الجزئي للمثال السابق.

<u>الحل:</u>

حصلنا على النتائج: $r_{y1} = 0.909$, $r_{y2} = 0.731$, $r_{12} = 0.741$

بتطبيق الصيغة الرياضية أعلاه لحساب معامل الارتباط الجزئي $r_{y2.1}$

$$r_{y21} = \frac{r_{y2} - r_{y1}\,r_{12}}{\sqrt{(1-r_{y1}^2)(1-r_{12}^2)}}$$

$$r_{y21} = \frac{0.931 - 0.909 \times 0.741}{\sqrt{(1-0.826)(1-0.549)}}$$

$$r_{y21} = \frac{0.257}{0.280}$$

$$r_{y21} = 0.92$$

إن قيمة معامل الارتباط الجزئي 0.92 تؤدي لمساهمة عالية (معامل التحديد الجزئي = 0.85) في تفسير تباين y ويمكن اختبار فرضية العدم بالاحصاءة t من أن معامل الارتباط الجزئي للمجتمع P تساوي صفر أي $H_o : P_{y1.2}$... k ونحسب t من الصيغة التالية

$$t = r_{y1.2}\sqrt{\frac{n-k-1}{1-r_{y1.2}^2}}$$

مع درجات حرية 1 – n – k = 6 – 3 – 1 = 2 ومقارنة t الناتجة بـ t الجدولية عند مستوى معنوية $\alpha = 0.05$.

ويكون الاختبار:

$H_o : P_{y1.2} = 0$ و $H_1 : P_{y1.2} \neq 0$ <u>ومن الجدول نجد أن:</u> $t_{2\,,\,0.025} = 4.303$

وبالتعويض:

$$t = r_{y1.2}\sqrt{\frac{n-k-1}{1-r_{y1.2}^2}}$$

$$t = 0.92\sqrt{\frac{2}{1-0.85}}$$

$$t = 0.92\sqrt{\frac{2}{0.15}}$$

$$t = 0.92 \times 3.65$$

$$t = 3.36$$

وحيث أن قيم t = 3.36 أقل من من قيمة t = 4.303 الجدولية فنقبل فرض العدم ويعني عدم معنوية معامل الارتباط الجزئي للمجتمع المتغير $P_{y2.1}$ X_1.

التباين المشترك :

يستخدم التباين المشترك في نظرية في الاحتمالات والإحصاء بين متغيرين عشوائيين لقياس انحرافاتهما المشتركة عن قيمهما المتوقعة . كما يُستخدم أيضًا لمجموعتين من البيانات العددية (الانحرافات عن المتوسطات). عادةً ما يكون التباين المشترك بين متغيرين عشوائيين مستقلين صفرا، مع أن العكس ليس صحيحًا دائما.

التباين المشترك هو امتداد لمفهوم التباين . الارتباط هو شكل معياري للتباين المشترك (بعد التباين المشترك بين متغيرين هو حاصل ضرب أبعادهما، بينما الارتباط هو كمية بلا أبعاد).

تعريف :

إن التباين المشترك لمتغيرين عشوائيين حقيقيين X و Y، لكل منهما تباين (محدود)، ويرمز له بـ Cov(X, Y)

أو أحيانًا σ_{XY} ، هو القيمة التالية:

$$\mathrm{Cov}(X,Y) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{X})(y_i - \bar{Y})$$

او:

$$\mathrm{Cov}(X,Y) \equiv \mathrm{E}[(X - \mathrm{E}[X])\,(Y - \mathrm{E}[Y])]$$

حيث تعبر $\mathbf{E}$ على الامل الرياضي، كما تجدر الإشارة إلى أن تباين المتغير X هو حالة خاصة من التباين المشترك، إذ يمكن التعبير عنه بصورة مكافئة على النحو التالي:

$$\mathrm{Var(X) = Cov(X, X)}$$

وفقًا لتعريفه، يُميّز التباين المشتركُ التغيراتِ المتزامنة لمتغيرَين عشوائيَّين: فهو يكون موجبًا حين تميل الانحرافات بين المتغيرات ومتوسطاتها إلى حمل الإشارة ذاتها، وسالبًا في الحالة المعاكسة:

- **موجب:**

الانحرافات $(x_i - \mu_X)$ و $(y_j - \mu_Y)$ تميل إلى حمل الإشارة ذاتها أي ارتفاع X يُصاحبه ارتفاع Y

- **سالب**

الانحرافات تميل إلى حمل إشارتَين متعاكستَين اي ارتفاع X يُصاحبه انخفاض Y

- **صفر**

حيث إن المتغيرَين غير مترابطَين؛ ومعامل ارتباطهما يساوي هو الآخر صفرًا. لكن: انعدام الترابط لا يعني الاستقلالية إلا في الحالة الطبيعية.

طبقًا لصيغة تعريفه، فإن بُعد التباين المشترك هو حاصل ضرب بُعدَي المتغيرَين. في المقابل، فإن معامل الارتباط، الذي يُعبَّر عنه بدلالة التباين والتباين المشترك، يتخذ قيمه في الفترة [−1، 1] ويبقى عديم البُعد (لا وحدةً له).

يتعلق بالوحدات $\dim[\mathrm{Cov}(X,Y)]=[X] \times [Y]$

عديم البُعد $r(X,Y)=\mathrm{Cov}(X,Y) / (\sigma_X \cdot \sigma_Y) \ni [-1, 1]$

حالة المتغيرات العشوائية المنفصلة

بالنسبة لمتغيرَين عشوائيَّين منفصلَين X وY يتخذان قيمهما على التوالي في مجموعتَين منتهيتَين:

$$\mathbf{X \in \{x_i \mid 1 \le i \le n\}} \text{ ، } \mathbf{Y \in \{y_j \mid 1 \le j \le m\}}$$

$$\mathrm{Cov}(X,Y)=\sum_{i=1}^{n}\sum_{j=1}^{m} x_i y_j \, \mathrm{P}(X=x_i \text{ et } Y=y_j)-\mathrm{E}[X]\,\mathrm{E}[Y]$$

حيث:

$$\sigma_X^2=\sum_{i=1}^{n} x_i^2 \, \mathrm{P}(X=x_i)-\mathrm{E}[X]^2 \qquad \sigma_Y^2=\sum_{j=1}^{m} y_j^2 \, \mathrm{P}(Y=y_j)-\mathrm{E}[Y]^2$$

ملاحظة: يُستخدم في الأدبيات العربية المصطلحُ التباين المشترك أو التغاير للدلالة على Covariance ، ومعامل الارتباط الخطي للدلالة على معامل Bravais-Pearson ، والاستقلالية للدلالة على Independence .

مصفوفة التباين المشترك:

تُعرَّف مصفوفة التباين المشترك لمتجه مكوَّن من p متغيرًا عشوائيًا: $\vec{X}=\begin{pmatrix} X_1 \\ \vdots \\ X_p \end{pmatrix}$ حيث يمتلك كلٌّ منها تباينًا ، بأنها المصفوفة المربعة التي حدُّها العام مُعطًى بـ:

$$a_{i,j}=\mathrm{Cov}(X_i, X_j)$$

$$\mathrm{Var}(\vec{X})=\begin{pmatrix} \mathrm{Var}(X_1) & \mathrm{Cov}(X_1,X_2) & \cdots & \mathrm{Cov}(X_1,X_p) \\ \mathrm{Cov}(X_2,X_1) & \ddots & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{Cov}(X_p,X_1) & \cdots & \cdots & \mathrm{Var}(X_p) \end{pmatrix}=\begin{pmatrix} \sigma_{x_1}^2 & \sigma_{x_1x_2} & \cdots & \sigma_{x_1x_p} \\ \sigma_{x_2x_1} & \ddots & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{x_px_1} & \cdots & \cdots & \sigma_{x_p}^2 \end{pmatrix}$$

تمارين محلولة :

الأسئلة النظرية

1. ماهي عوامل التحكم في دقة معامل الارتباط الخطي؟
2. متى نستعمل معامل ارتباط سبيرمان (Spearman)؟
3. ما هي الشروط التي يجب ان تتوفر في المتغيرات من اجل استعمال معامل الارتباط لبيرسون؟

التمرين الاول

قمنا بإجراء دراسة حول عينة مكونة من 30 فرد لتحديد إذا كان الضغط الدموي للفرد يتأثر بالسن، النتائج الإحصائية الوصفية هي ممثلة في الجدول التالي:

	السن	الضغط الدموي
المتوسط الحسابي	45	143
الانحراف المعياري	15	23
مجموع (السن * الضغط الدموي)	199576	

المطلوب:

1. حدد الاجابات الصحيحة، مع تصحيح الخاطئة:

1) في هده الدراسة:

أ- السن هو متغير نوعي

ب- حجم العينة هو 30

ج- الضغط الدموي هو متغير كمي مستمر

د- تباين السن يساوي $\sqrt{15}$

هـ- الاجابات السابقة خاطئة

2) لتحديد هل هناك علاقة بين السن والضغط الدموي، يمكن استعمال:

أ- اختبار t لمعرفة الفروق بين المتوسطات

ب- اختبار كاي تربيع

ج- اختبار معامل الارتباط

د- معامل الارتباط الرتبي

هـ- الاجابات السابقة خاطئة

3) قبل تقدير النموذج الخطي الدي يمثل العلاقة بين السن والضغط الدموي يجب التأكد ممايلي:

أ- درجة المعنوية $P > 0.05$

ب- عدم وجود علاقة بين المشاهدات

ج- علاقة خطية بين المتغيرين

د- تكرار المتوقع بالنسبة للفرضية H_0 يكون دائما اكبر او يساوي 5

هـ- الاجابات السابقة خاطئة

2. ماهي طبيعة الإشكالية بين المتغيرين؟
3. حدد العلاقة بين المتغيرين عند درجة معنوية 5%، (تعطى $t_{tab} = 2.045$)

التمرين الثاني :

قامت لجنة مراقبة في مذبحة ما بدراسة القياسات الخاصة بالذبيحة، وكانت النتائج ملخصة في الجدول كما يلي:

القياسات	X1	X2	X3
1	224	79.1	3.00
2	232	73.4	8.70
3	233	76.4	7.00
4	240	75.3	8.70
5	217	76.5	7.80
6	243	77.4	7.10
7	229	78.4	4.60
8	240	76.5	8.20

حيث ان: X_1 : وزن الذبيحة الكلي ، X_2 : وزن اللحم ، X_3 : وزن الدهون.

المطلوب:

1. احسب مصفوفة الارتباط بين المتغيرات الظاهرة.
2. احسب معامل الارتباط المتعدد ومعامل الارتباط الجزئي مع التعليق على النتائج.
3. ما هو الفرق بين معامل الارتباط الجزئي والمتعدد؟

الحل :

1. ماهي عوامل التحكم في دقة معامل الارتباط الخطي؟

- أن تقع نقاط الأزواج (x , y) على خط مستقيم أو تكون قريبة جداً منه حتى تحقق صفة أن العلاقة خطية ($y = ax + b$) ويمكن ملاحظة ذلك من شكل الانتشار. إن لم تكن العلاقة خطية فستخدم معامل آخر.
- مقدار التباين فالعلاقة طردية بين الزيادة في التباين ومعامل الارتباط.
- دقة معامل الارتباط تتأثر بحجم العينة.

- شكل التوزيع وتماثله للمتغيرين يزيد من قيمة معامل الارتباط فإن كان شكلا التوزيع متماثلين فيكون $r = \pm 1$ وإن كانا الالتواء في نفس الاتجاه كان $r = 1$ وإن كان الالتواء في اتجاهين متضادين (احدهم التواءه موجب والآخر سالب) كان $r = -1$

2. متى نستعمل معامل ارتباط سبيرمان (Spearman) ؟

نستعمل معامل الارتباط للرتب (سبيرمان) في حالة البيانات الوصفية (النوعية) التي يمكن ترتيبها.

التمرين الاول:

قمنا بإجراء دراسة حول عينة مكونة من 30 فرد لتحديد إذا كان الضغط الدموي للفرد يتأثر بالسن، النتائج الإحصائية الوصفية هي ممثلة في الجدول التالي:

	السن	الضغط الدموي
المتوسط الحسابي	45	143
الانحراف المعياري	15	23
مجموع (السن * الضغط الدموي)	199576	

المطلوب:

4. حدد الاجابات الصحيحة، مع تصحيح الخاطئة:

4) في هده الدراسة:

أ- السن هو متغير نوعي (متغير كمي)

ب- حجم العينة هو 30

ج- الضغط الدموي هو متغير كمي مستمر (متغير كمي متقطع)

د- تباين السن يساوي $\sqrt{15}$ ($var = 15^2$)

هـ- الاجابات السابقة خاطئة

أ- اختبار t لمعرفة الفروق بين المتوسطات

ب- اختبار كاي تربيع (المتغيرات ليست نوعية)

ج- اختبار معامل الارتباط

د- معامل الارتباط الرتبي

هـ- الاجابات السابقة خاطئة

5) لتحديد هل هناك علاقة بين السن والضغط الدموي، يمكن استعمال:

6) قبل تقدير النموذج الخطي الدي يمثل العلاقة بين السن والضغط الدموي يجب التأكد مما يلي:

أ- درجة المعنوية $P > 0.05$

ب- عدم وجود علاقة بين المشاهدات

ج- علاقة خطية بين المتغيرين

د- تكرار المتوقع بالنسبة للفرضية H_0 يكون دائما اكبر او يساوي 5

هـ- الاجابات السابقة خاطئة

5. ماهي طبيعة الإشكالية بين المتغيرين؟

هل هناك علاقة بين السن والضغط الدموي (علاقة تناظرية)

هل يؤثر السن على الضغط الدموي (علاقة سببية)

6. حدد العلاقة بين المتغيرين عند درجة معنوية

$$r = \frac{\operatorname{cov}(X,Y)}{\sqrt{s_X^2 \times s_Y^2}} \Longrightarrow \hat{\operatorname{cov}}(X,Y) = \frac{\sum x_i y_i - \frac{(\sum x_i)(\sum y_i)}{n}}{(n-1)}$$

$$r = \frac{\frac{\sum x_i y_i}{n} - \bar{x}\bar{y}}{\sigma_x \sigma_y} = \frac{\frac{199576}{30} - 45 * 143}{15 * 23}$$

علاقة متوسطة طردية $= 0.63$

الاختبار هنا: $H_0: \rho = 0$ أي لا توجد علاقة.

$H_1: \rho \neq 0$ أي توجد علاقة. إحصاء الاختبار المتبع هنا هو

$$T = r\sqrt{\frac{n-2}{1-r^2}} \qquad T = \sqrt{\frac{30-2}{1-0.63^2}} = 6.81$$

وحيث أن $T = 6.81$ وهي أكبر من 2.045 فنرفض H_0 ونقبل H_1 (الفرضية البديلة) أي توجد علاقة.

القياسات	X_1	X_2	X_3	X_1^2	X_2^2	X_3^2	X_1*X_2	X_1*X_3	X_2*X_3
1	224	79,1	3	50176	6256,81	9	17718,4	672	237,3

2	232	73,4	8,7	53824	5387,56	75,69	17028,8	2018,4	638,58
3	233	76,4	7	54289	5836,96	49	17801,2	1631	534,8
4	240	75,3	8,7	57600	5670,09	75,69	18072	2088	655,11
5	217	76,5	7,8	47089	5852,25	60,84	16600,5	1692,6	596,7
6	243	77,4	7,1	59049	5990,76	50,41	18808,2	1725,3	549,54
7	229	78,4	4,6	52441	6146,56	21,16	17953,6	1053,4	360,64
8	240	76,5	8,2	57600	5852,25	67,24	18360	1968	627,3
المجموع	1858	613	55,1	432068	46993,24	409,03	142342,7	12848,7	4199,97

التمرين الثاني

1- مصفوفة الارتباط بين المتغيرات الظاهرة:

معامل الارتباط المتعدد

$$r_{y1} = \frac{n\sum yX_1 - \sum y \sum X_1}{\sqrt{n\sum y^2 - (\sum y)^2}\sqrt{n\sum X_1^2 - (\sum X_1)^2}} =$$

$$A = \begin{array}{c|ccc} & X1 & X2 & X3 \\ \hline X1 & 1 & & \\ X2 & -0,24 & 1 & \\ X3 & 0,41 & -0,86 & 1 \end{array}$$

2- احسب معامل الارتباط المتعدد:

$$R_{123} = \sqrt{\frac{r_{12}^2 + r_{13}^2 - 2r_{12}r_{13}r_{23}}{1 - r_{23}^2}} \Longrightarrow R_{x1x2x3} = 0.46$$

الارتباط المتعدد بين المتغيرات الثلاثة هو طردي متوسط

3- معامل الارتباط الجزئي:

$$r_{y2.1} = \frac{r_{y2} - r_{y1}\, r_{12}}{\sqrt{(1 - r_{y1}^2)(1 - r_{12}^2)}} \Longrightarrow R_{x1x2\,;\,x3} = 0.24$$

الارتباط بين المتغيرين X_1 و X_2 هو طردي ضعيف، و ذلك مع تثبيت اثر المتغير X_3

3- معامل الارتباط المتعدد يقيس قوة العلاقة بين أكثر من متغيرين وهي مغيرات عشوائية متصلة التوزيع (توزيع متعدد المتغيرات Multivariate distribution) وإن حساب قيمة R هو امتداد لقيمة معامل الارتباط البسيط.

معامل الارتباط الجزئي يقيس الارتباط بين أي زوج من المتغيرات عند ثبات المتغيرات الأخرى، ويستخدم لتحديد العلاقة بين متغيرين محددين أو للتعرف على متغيرات يراد حذفها بسبب عدم تأثيرها على المتغير التابع.

تمارين

تمرين 1:

تقدم عشرة طلاب لامتحان المرحلة الثانوية وكانت معدلات نتائجهم حسب الصف والمدرسة كالتالي والمطلوب حساب معامل سبيرمان للارتباط.

معدل الطالب في الصف (X)	73	80	70	66	88	71	65	88	92	74
مدل الطالب في المدرسة (Y)	64	78	64	70	92	64	55	90	88	72

تمرين 2:

اختيرت عينة عشوائية مكونة من 10 طلاب في لعبتي كرة اليد وكرة القدم وسجلت نتائجهم في الجدول الآتي والمطلوب معرفة العلاقة بين أداء الطالب في اللعبتين.

أداء الطالب في كرة اليد(X)	جيد	ضعيف	جيد	ممتاز	جيد جداً	جيد	ممتاز	مقبول	جيد جداً	ضعيف
أداء الطالب في كرة القدم(Y)	ضعيف	مقبول	مقبول	جيد	جيد	جيد جداً	ممتاز	جيد	ضعيف	جيد

تمرين 3:

الجدول التالي يبين مشاهدات 150 مربية أطفال حسب مستواهم العلمي في جودة التكفل بالأطفال والمطلوب قياس مدى اختلاف التكرارات المشاهدة والتكرارات المتوقعة عند مستوى معنوية 0.01 (α=0.01) باستخدام معامل كا^2(χ^2).

المستوى التعليمي	الملاحظة

	كثيراً جداً	كثيراً	متوسط	قليلاً	قليلاً جداً
شهادة الثانوية أوما يعادلها	25	17	11	4	3
دبلوم سنتين بعد الثانوية	22	15	7	2	2
الشهادة الجامعية	21	12	6	3	0

تمرين 4:

- تم تحليل نقاط الامتحان السداسي الأول في مادتي الاقتصاد القياسي والرياضيات لمجموعة من الطلبة قدر عددهم بـ 60 طالب، نريد معرفة علاقة بين نقاط الطلبة في مادة الاقتصاد القياسي والرياضيات ، نتائج الإحصاء الوصفي لهذا التحليل هي ملخصة في الجدول الموالي:

	الاقتصاد القياسي	الرياضيات
المتوسط الحسابي	13.2	12.7
الانحراف المعياري	1.5	2.6
مجموع (اقتصاد قياسي * رياضيات)	10173.00	

المطلوب :

1- ما هي الإشكالية التي سنجيب علها في هذا التمرين؟

2- حدد فرضيات الاختبار الاحصائي لهذه الإشكالية؟

3- ما هو الاختبار الاحصائي الذي سنستعمله؟ ولماذا؟

4- ما هي شروط قبول هذا الاختبار؟

5- قم بحساب قيمة هذا الاختبار الاحصائي.

6- ماذا تستنتج عند مستوى معنوية 0.05؟

تمرين 5

نستعمل المعطيات الواردة في الجدول التالي لدراسة علاقة رقم الأعمال (y) لـ 05 مؤسسة تجارية من نفس القطاع بقيمة رأس المال الثابت (x_1) والمتداول (x_2) الذي تستعمله هذه المؤسسات في نشاطها. (القيم معبر عنها بألاف الدولارات)

نفترض أن نموذج الانحدار المعبر عنه في هذه العلاقة هو خطي متعدد.

القيمة السنوية المتوسطة لرأس المال المتداول (x_2)	القيمة السنوية المتوسطة لرأس المال الثابت (x_1)	القيمة السنوية لرقم الأعمال (y)	رقم المؤسسة
105	118	203	**1**
56	28	63	**2**
54	17	45	**3**
63	50	113	**4**
28	56	121	**5**

المطلوب :

1. احسب معامل الارتباط الجزئي $r_{yx1,x2}$ و معامل الارتباط المتعدد، علق على النتائج، ما الفرق بينهما؟
2. حدد مصفوفة الارتباط و معنوية معاملات الارتباط مع التعليق على النتائج (تعطى قيم ستودنت المحسوبة اسفل الورقة)
3. ما هو التحليل العاملي المناسب لهذه المعطيات، علل؟
4. مثل بيانيا توزع المؤسسات بدلالة المتغيرات y، X_1، X_2

الفصل الرابع : التحليل الانحدار المتعدد

الفصل الرابع : التحليل الانحدار المتعدد

تهيد:

يُعد الانحدار المتعدد امتدادا منهجيا للانحدار الخطي البسيط، يهدف إلى نمذجة العلاقة بين متغير تابع واحد ومتغيرين مستقلين أو أكثر في آن واحد، مما يتيح عزل الأثر الجزئي لكل متنبئ مع ثبات العوامل الأخرى ((Ceteris Paribus، وهو ما يعكس طبيعة الظواهر الواقعية المعقدة التي نادراً ما تتأثر بعامل وحيد؛ ويعتمد النموذج على معادلة خطية تجمع بين ثابت قاطع ومعاملات انحدار خاصة بكل متغير مستقل بالإضافة إلى حد خطأ عشوائي، ويتطلب تطبيقه السليم التحقق من افتراضات إحصائية جوهرية مثل خطية العلاقة، استقلالية البواقي، تجانس التباين، والتوزيع الطبيعي للأخطاء، مع مراقبة دقيقة لمشكلتي التعدد الخطي والإفراط في التخصيص، مما يجعله أداة تحليلية حيوية للتنبؤ الدقيق والتفسير المعياري في مجالات الاقتصاد، العلوم الاجتماعية، الطب، وهندسة البيانات.

في هذا الفصل، سنعمم ونوسع النتائج الانحدار البسيط التي تم دراستها في مادة الاقتصاد القياسي للحالة الأكثر أهمية والتي نهدف فيها إلى شرح متغير تابع Y من خلال مجموعة من المتغيرات المستقلة X. ولتبسيط الرموز والمعادلات، سنعتمد على التدوين المصفوفي. ويُعد نموذج الانحدار الخطي المتعدد الأداة الإحصائية الأكثر شيوعاً واستخداماً في دراسة البيانات متعددة الأبعاد، وباعتباره حالة خاصة من النماذج الخطية العامة، فإنه يمثل الامتداد الطبيعي للانحدار الخطي البسيط.

1. الصياغة العامة لنموذج الانحدار المتعدد:

يمكن صياغة النموذج الانحدار الخطي المتعدد كالتالي:

$$p = 2 \;:\; Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i \qquad i = 1:n$$

$$p > 2 \;:\; Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_p X_{ip} + \varepsilon_i \qquad i = 1:n$$

حيث :

X_{ij} : هو الـ $j_{éme}$ متغير مفسر – مستقل للفرد i ‎(j= 1 :p)

ε_i : حد الخطأ العشوائي ، مستقل عن x ويتبع قانون طبيعي $\varepsilon \sim N(0, \sigma_\varepsilon^2 I)$

$\beta_0, \beta_1, \ldots, \beta_p$: هي المعالم المجهولة للنموذج.

كما يمكن صياغة النموذج على الشكل المصفوفي التالي :

$$Y = XB + \varepsilon$$

$$Y = \begin{pmatrix} Y_1 \\ . \\ . \\ . \\ Y_n \end{pmatrix}; \qquad X = \begin{pmatrix} 1 & X_{11} & X_{12} & \ldots & X_{1p} \\ . & . & & & . \\ . & . & & & . \\ . & . & & & . \\ 1 & X_{n1} & X_{n2} & \ldots & X_{np} \end{pmatrix}$$

$$B = \begin{pmatrix} \beta_0 \\ . \\ . \\ . \\ \beta_p \end{pmatrix}; \qquad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ . \\ . \\ . \\ \varepsilon_n \end{pmatrix}$$

حيث:

$Y(n \times 1)$: المتغير التابع او المفسر

$X(n \times (k+1))$ مصفوفة المتغيرات المستقلة او المفسرة

$\beta((k+1) \times 1)$ شعاع المعالم

$\varepsilon(n \times 1)$ شعاع الأخطاء

2.الفرضيات الأساسية للنموذج:

كما هو الحال في الانحدار البسيط، تتيح الفرضيات تحديد: خصائص المقدرات (كالانحياز والتقارب)؛ وتوزيعاتها الاحتمالية (المستخدمة في التقدير بفترات الثقة واختبارات الفرضيات).

توجد فئتان رئيسيتان من الفرضيات:

- **الفرضيات العشوائية(Stochastic Hypotheses) :**

1) H_1 : المتغيرات المستقلة X_j محددة دون خطأ، حيث : $j = 1, \ldots, p$؛
2) H_2 : $\mathbf{E(\varepsilon_i) = 0}$ أي النموذج مُحدد بشكل صحيح في المتوسط؛
3) H_3 : $\mathbf{Var(\varepsilon_i) = \sigma^2 \;\; \forall \;\; i}$ أي تجانس تباين الأخطاء (ثبات التباين)؛
4) H_4 : $\mathbf{cov(\varepsilon_i, \varepsilon_j) = 0 \;\; \forall \;\; i \neq j}$ أي عدم وجود ارتباط ذاتي بين الأخطاء؛
5) H_5 : $\mathbf{cov(X_i, \varepsilon_j) = 0 \; \forall \; i \neq j}$ أي استقلالية الأخطاء خطياً عن المتغيرات الخارجية؛
6) H_6 : $\mathbf{\varepsilon \sim N_n(0, \sigma^2 I_n)}$ أي تتبع الأخطاء توزيعاً طبيعياً متعدد الأبعاد.

ملاحظة *: تستلزم H_6 الفرضيات H_2 و H_3 وH_4، لكن العكس غير صحيح؛ فاجتماع الفرضيات الثلاث لا يضمن بالضرورة أن ε يتبع توزيعاً طبيعياً متعدد الأبعاد*

يمكن كتابة الفرضية H_6 على الشكل المصفوفي التالي :

$$\mathrm{H}_2 : \mathbb{E}(\varepsilon) = \mathbb{E}\begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}$$

- **الفرضيات الهيكلية(Structural Hypotheses)**

1) H_7 : عدم وجود تعدد خطي بين المتغيرات التوضيحية، أي أن المصفوفة $X^T X$ غير شاذة (منتظمة)، حيث $\mathbf{det(X^T X) \neq 0}$ وتوجد المصفوفة العكسية $(X^T X)^{-1}$

ملاحظة: هذا يعادل شرط الرتبة $(rank(X) = rank(X^T X) = p + 1)$؛

2) H_8 : $\mathbf{(1/n) X^T X}$ تتقارب نحو مصفوفة محدودة غير شاذة **Q** عندما $n \longrightarrow +\infty$ ؛
3) H_9 : $\mathbf{n > p + 1}$ اي عدد المشاهدات أكبر تماماً من عدد المتغيرات زائد الثابت. فلو تساوى العددان، لساوى عدد المعادلات عدد المجاهيل **aj**، ولمر خط الانحدار بجميع النقاط، مما يحول المشكلة إلى حالة استيفاء خطي problème d'interpolation linéaire.

في ظل فرضيتي تجانس التباين وعدم الارتباط الذاتي، يمكن كتابة مصفوفة التباين والتغاير لمتجه الأخطاء على النحو التالي:

$$\mathrm{H}_3 \text{ et } \mathrm{H}_4 : \operatorname{cov}(\varepsilon) = \sigma^2 I_n = \sigma^2 \underbrace{\begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & \cdots & 1 \end{pmatrix}} = \begin{pmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & \cdots & \sigma^2 \end{pmatrix}$$

3. تقدير شعاع المعالم β :

- **طريقة المربعات الصغرى (Estimateur des MCO) :**

من النموذج الأصلي : $y_i = b_0 + b_1 x_{i,1} + \cdots + b_p x_{i,p} + \varepsilon_i$

سنقوم بتقدير المعالم لنتحصل على النموذج التالي:

$$\hat{y}_i = \hat{b}_0 + \hat{b}_1 x_{i,1} + \cdots + \hat{b}_p x_{i,p}$$

الأخطاء العشوائية او البواقي المقدرة ينتجان عن الفرق بين القيمتين y و $\hat{y}$ أي :

$$\hat{\epsilon}_i \equiv y_i - \hat{y}_i$$

تُقدر المعاملات بإيجاد الحل الذي يُقلّل مجموع مربعات البواقي:

$$\min \sum_{i=1}^{n} \hat{\varepsilon}_i^2 = \min_{\hat{b}_0,\ldots,\hat{b}_p} \sum_{i=1}^{n} \left(y_i - \hat{b}_0 - \hat{b}_1 x_{i,1} - \cdots - \hat{b}_p x_{i,p}\right)^2$$

يسمى مُقدر المربعات الصغرى العادية ، ويُرمز له بـ (MCO) للمعلمة β ، والمُشار إليه بـ $\hat{\beta}$ ، بالقيمة التالية:

$$\hat{\beta} = \underset{\beta_1,\ldots\beta_p}{\arg\min} \sum_{i=1}^{n} \left(y_i - \sum_{j=1}^{p} \beta_j x_{ij}\right)^2 = \underset{\beta\in\mathbb{R}^p}{\arg\min} \left(Y - X\beta\right)' \left(Y - X\beta\right)$$

إذا تحققت الفرضية H_1 ، فإن مُقدر المربعات الصغرى العادية (MCO) للمعلمة β ، والمُشار إليه بـ $\hat{\beta}$ ، يُعطى بالصيغة المصفوفية التالية:

$$\hat{\beta} = (X^\top X)^{-1} X^\top Y$$

هو المُقدِّر الذي يُقلّل مجموع مربعات البواقي

ملاحظة : لماذا نُقلّل مجموع المربعات لا المجموع البسيط؟ لأن متوسط البواقي يساوي صفراً، مما يجعل المجموع البسيط يلغي نفسه (بواقٍ موجبة وسالبة). المربعات تتفادى هذه المشكلة.

- **خصائص مصفوفة X^TX :**
  - إذا كانت X_j مُمركزة : $\frac{1}{n} X^T X$ هي مصفوفة التباين المشترك للمتغيرات الخارجية.
  - إذا كانت X_j مُمركزة ومُوحَّدة(centrées-réduites) : $\frac{1}{n} X^T X$ هي مصفوفة الارتباط.

- **التفسيرات الهندسية والجبرية والإحصائية لمُقدِّر MCO**

1- هندسياً - الإسقاط المتعامد : يُقابل مُقدِّر MCO الإسقاط المتعامد للمتجه Y على الفضاء المُتوسَّع بمتجهات X :

$$\hat{Y} = X(X^TX)^{-1}X^TY = P_XY$$

2- جبرياً - المعكوسة المعممة: بضرب النظام $Y = Xb$ يساراً بالمعكوسة المعممة $(X^TX)^{-1}X^T$:

$$(X^TX)^{-1}X^TY = (X^TX)^{-1}X^TX\,b = b$$

$$\therefore \quad \hat{b} = (X^TX)^{-1}X^TY$$

3- إحصائياً - مبدأ الإمكان الأعظم: تحت فرضية التوزيع الطبيعي للأخطاء $\varepsilon \sim \mathcal{N}(0, \sigma^2 I)$ ، يتطابق مُقدِّر MCO مع مُقدِّر الإمكان الأعظم :

$$\hat{b}_{MCO} \equiv \hat{b}_{MV}$$

4. خصائص المُقدِّرات:

➢ الخصائص في العينات المحدودة:

1- عدم التحيز - Estimateur sans biais : $E(\hat{\beta}) = \beta$ تحت الفرضيات H_1, H_2, H_5. هذه الخاصية لا تتأثر بوجود ارتباط ذاتي أو عدم تجانس التباين.

2- **أفضل مُقدِّر خطي غير متحيز - BLUE (Gauss-Markov) :** تحت H_1 إلىH_5 ، لا يوجد مُقدِّر خطي غير متحيز لـ β بتباين أصغر.

3- **التوزيع الطبيعي:** $\hat{\boldsymbol{\beta}} \sim N(\boldsymbol{\beta}, \boldsymbol{\sigma}^2(X^TX)^{-1})$ تحت H_1, H_2, H_6

➢ الخصائص التقاربية:

4- الاتساق بالاحتمال: $\hat{\beta} \xrightarrow{p} a$ عندما $n \to \infty$ تحت H_6 وH_8

5- التوزيع التقاربي الطبيعي: تحت H_1 إلى H_5 وH_8، دون الحاجة إلى فرضية H_6 (التوزيع الطبيعي للأخطاء).

5. تقييم النموذج

➢ مصفوفة التباين-التغاير للمعاملات

تُعدّ ضرورية لتحديد تباين كل معامل مُقدَّر وإجراء اختبارات الفرضيات:

$$\mathbf{Var(\hat{a}) = \boldsymbol{\sigma}^2(X^TX)^{-1}}$$

يُقدَّر تباين البواقي بالمُقدِّر غير المتحيز:

$$\mathbf{\hat{\sigma}^2 = SCR / (n - p - 1) = \Sigma \hat{e}_i^2 / (n - p - 1)}$$

حيث $(n-p-1)$ = درجات الحرية (عدد المشاهدات – عدد المعاملات المُقدَّرة)

عند وجود عدم تجانس التباين أو ارتباط ذاتي، تُصحَّح المصفوفة بـ:

- **مصفوفة و ايت(HC)** : متسقة في حالة عدم تجانس التباين
- **مصفوفة نيوي-ويست(HAC)** : متسقة في حالتَي عدم تجانس التباين والارتباط الذاتي معاً.

➢ **توزيع المعاملات — اختبار ستيودنت**

الإحصاء التالي يتبع قانون ستيودنت بـ $(n-p-1)$ درجة من الحرية:

$$t = (\hat{a}_j - a_j) / \hat{s}(\hat{a}_j) \sim t(n-p-1)$$

اختبار نفي المعامل: $H_0 : a_j = 0$ مقابل $H_1 : a_j \neq 0$ يُحدد ما إذا كان المتغير x_j يؤدي دوراً معنوياً في النموذج.

ملاحظة : قبول الفرضية الصفرية قد يدل على غياب الارتباط الحقيقي، أويمكن أن ينجم عن الارتباط الشديد بين x_j ومتغير خارجي آخر (مشكلة التعددية الخطية.(multicolinéarité)

6. جدول تحليل التباين

يرتكز التقييم الشامل على معادلة تحليل التباين:

$$SCT = SCE + SCR$$

- **SCT**: مجموع المربعات الكلي: التباين الكلي للمتغير التابع
- **SCE**: مجموع المربعات المُفسَّر: التباين المُفسَّر بالنموذج
- **SCR**: مجموع المربعات البقايا: التباين غير المُفسَّر

مصدر التباين	مجموع المربعات	درجات الحرية	المتوسطات التربيعية
مُفسَّر	SCE	p	SCE / p
بواقي	SCR	n − p − 1	SCR / (n−p−1)
الإجمالي	**SCT**	**n − 1**	—

➢ **معامل التحديد R^2**

$$R^2 = SCE / SCT = 1 - (SCR / SCT)$$

R^2 يتراوح بين 0 و1. كلما اقترب من 1 كان النموذج أكثر أهمية. لكنه يميل إلى الارتفاع ميكانيكياً مع زيادة عدد المتغيرات، لذا يُوصى عند مقارنة النماذج باستخدام:

$$R^2\_adj = 1 - [(1 - R^2)(n - 1)] / (n - p - 1)$$

R^2 المُعدَّل دائماً أقل من R^2 الأصلي.

- **الدلالة الإجمالية - اختبار فيشر**

- **H_0 :** $a_1 = a_2 = ... = a_p = 0$ (لا يوجد أي متغير له أثر معنوي)
- **H_1 :** على الأقل أحد المعاملات غير مساوٍ للصفر

$$F = [R^2 / p] / [(1 - R^2) / (n - p - 1)] \sim F(p, n-p-1)$$

تُرفض H_0 إذا كانت $F_{calc} > F_{(1-\alpha)(p, n-p-1)}$ ، أو إذا كانت p-value أقل من مستوى المعنوية α.

مثال :

نريد تفسير رقم المبيعات Y بالمليون دج) لشركة ما انطلاقاً من متغيرين:

X_1 : ميزانية الإشهار (بالمليون دج)

X_2 :عدد نقاط البيع

المشاهدة i	Y (المبيعات)	X_1 (الإشهار)	X_2 (نقاط البيع)
1	10	2	3
2	15	3	4
3	12	2	5
4	20	4	6
5	18	4	5
6	25	5	7
المتوسط	**16.67**	**3.33**	**5.00**

1. النموذج المراد تقديره:

$$y_i = b_0 + b_1 x_{i,1} + b_2 x_{i,2} + \varepsilon_i$$

2. الصياغة المصفوفية: نبني مصفوفة X بإضافة عمود الواحدات للثابت b_0 :

$$y = \begin{pmatrix} 10 \\ 15 \\ 12 \\ 20 \\ 18 \\ 25 \end{pmatrix}, \quad X = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 4 \\ 1 & 2 & 5 \\ 1 & 4 & 6 \\ 1 & 4 & 5 \\ 1 & 5 & 7 \end{pmatrix}, \quad b = \begin{pmatrix} b_0 \\ b_1 \\ b_2 \end{pmatrix}$$

3. تقدير المعاملات بطريقةMCO :

$$\hat{b} = (X^T X)^{-1} X^T y$$

- **حساب X^TX :**

$$X^T X = \begin{pmatrix} 6 & 20 & 30 \\ 20 & 74 & 104 \\ 30 & 104 & 150 \end{pmatrix}$$

- حساب X^Ty :

$$X^T y = \begin{pmatrix} 100 \\ 349 \\ 518 \end{pmatrix}$$

- المعاملات المُقدَّرة:

$$\hat{b} = (X^T X)^{-1} X^T y \approx \begin{pmatrix} 1.02 \\ 2.58 \\ 1.37 \end{pmatrix}$$

- معادلة الانحدار المُقدَّرة:

$$\hat{y}_i = 1.02 + 2.58\, x_{i,1} + 1.37\, x_{i,2}$$

- التفسير:

كل زيادة بمليون دج في الإشهار ترفع المبيعات بـ 2.58مليون دج (مع ثباتX_2)

كل نقطة بيع إضافية ترفع المبيعات بـ 1.37مليون دج (مع ثباتX_1)

4. جودة النموذج - R^2 واختبار فيشر:

- القيم المُقدَّرة والبواقي:

i	y_i	$\hat{y}_i$	$\hat{e}_i = y_i - \hat{y}_i$	$\hat{e}_i^2$
1	10	10.25	−0.25	0.06
2	15	15.20	−0.20	0.04
3	12	12.58	−0.58	0.34
4	20	20.13	−0.13	0.02
5	18	18.76	−0.76	0.58
6	25	24.71	0.29	0.08
المجموع			≈ 0	SCR ≈ 1.12

- حساب SCT وSCE :

$$SCT = \sum_{i=1}^{6} (y_i - \bar{y})^2 = 158.33$$

$$SCE = SCT - SCR = 158.33 - 1.12 = 157.21$$

$$R^2 = \frac{SCE}{SCT} = \frac{157.21}{158.33} \approx 0.993$$

$$F = \frac{SCE/p}{SCR/(n-p-1)} = \frac{157.21/2}{1.12/3} = \frac{78.60}{0.373} \approx 210.7$$

➢ **الاستنتاج:**

$R^2 = 0.993$: النموذج يُفسّر 99.3% من تباين المبيعات

$F = 210.7 >> F\_critique(2,3) = 9.55$ عند $\alpha=5\%$ ⟶ النموذج دال إحصائياً

➢ **التنبؤ بقيمة جديدة**

إذا أرادت الشركة رفع ميزانية الإشهار إلى 6 مليون دج مع 8 نقاط بيع:

$$\hat{y}_{new} = 1.02 + 2.58 \times 6 + 1.37 \times 8 = 1.02 + 15.48 + 10.96 = \mathbf{27.46}$$

تمارين :

التمرين 1:

قامت لجنة مراقبة في مذبحة ما بدراسة القياسات الخاصة بالذبيحة، وكانت النتائج ملخصة في الجدول كما يلي:

حيث ان: Y : وزن الذبيحة الكلي ، X_1 : وزن اللحم ، X_2 : وزن الدهون.

القياسات	Y	X2	X3
1	**224**	**79.1**	**3.00**
2	**232**	**73.4**	**8.70**
3	**233**	**76.4**	**7.00**
4	**240**	**75.3**	**8.70**
5	**217**	**76.5**	**7.80**
6	**243**	**77.4**	**7.10**

<u>المطلوب:</u>

1. احسب مصفوفة الارتباط بين المتغيرات الظاهرة. مع التعليق على النتائج
2. قم بتقدير معادلة الانحدار الخطي التي توضح العلاقة بين وزن الذبيحة الكلي ووزن اللحم.
3. علق على النتائج
4. ما هي اهم فرضيات طريقة المربعات الصغرى باختصار؟

<u>تمرين 2:</u>

الجدول يوضح معدل الدخل السنوي x ومقدار التوفير السنوي (بالدينار) y لعينة تتكون من 10 أسر:

الدخل السنوي x1	12	15	9	22	16	36	10	11	25	18
الاستهلاك السنوي x2	4	6.3	5.8	7.3	6.1	6.6	5.9	6.4	7	6.3

التوفير السنوي y	0.6	1.1	0.2	2.4	1.2	3.6	0.4	0.6	1.7	1.1

المطلوب:

1. كتابة النموذج الخطي؟
2. تقدير معادلة الانحدار الخطية باستخدام طريقة المربعات الصغرى مستخدما معدل التوفير السنوي كمتغير تابع ؟
3. حساب تباين البواقي؟
4. حساب تباين المقدرات؟
5. إيجاد مجال الثقة لمعالم النموذج؟
6. حساب معامل التحديد؟
7. إيجاد جدول تحميل التباين؟
8. اختبار المعنوية الإحصائية لمعالم النموذج عند مستوى معنوية $\alpha = 5\ \%$ ؟
9. اختبار المعنوية الكمية للنموذج عند مستوى معنوية $\alpha = 5\ \%$ ؟

تمرين 3:

يمثل الجدول أدناه تطور الدخل المتاح الإجمالي والاستهلاك الأسري باليورو لدولة معينة خلال الفترة 1992-2001:

السنة	الدخل	الاستهلاك
1992	8000	7389.99
1993	9000	8169.65
1994	9500	8831.71
1995	9500	8652.84
1996	9800	8788.08
1997	11000	9616.21
1998	12000	10593.45
1999	13000	11186.11
2000	15000	12758.09
2001	16000	13869.62

نهدف إلى تفسير استهلاك الأسر (C) من خلال الدخل (R) وفقًا للمعادلة التالية:

$$C_t = \alpha + \beta R_t + \varepsilon_t$$

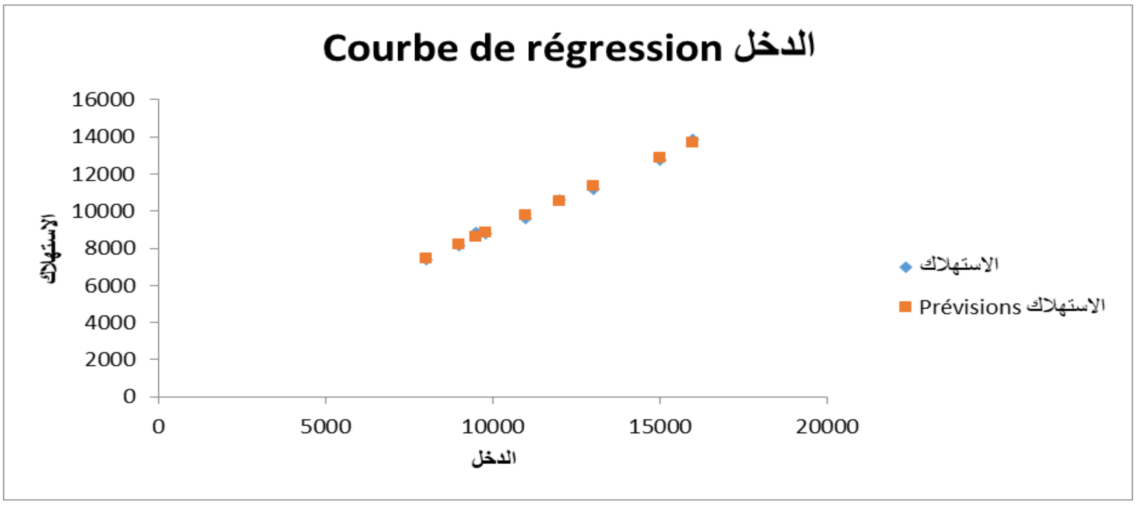
المطلوب:

1. رسم سحابة النقاط والتعليق عليها. حساب معامل الارتباط الخطي بين سلسلتي الاستهلاك والدخل.
2. حساب مقدرات المربعات الصغرى للمعاملين $\hat{\alpha}$ و $\hat{\beta}$
3. استنتاج القيم المقدرة $\hat{C}_t$ للاستهلاك.
4. حساب البواقي والتحقق من خاصية أن متوسط البواقي يساوي صفرًا.
5. حساب تقدير التباين للخطأ، الذي يرمز له بـ $\hat{\sigma}_\varepsilon^2$
6. هل المعامل β ، الذي يمثل الميل الحدي للاستهلاك، مختلف معنويًا عن الصفر؟
7. حدد مجال الثقة بنسبة 95% لهذا المعامل.
8. حساب معامل التحديد وإجراء اختبار فيشر لتحديد ما إذا كانت معادلة الانحدار ذات دلالة إحصائية ككل.
9. كتابة والتحقق من معادلة تحليل التباين، ثم تفسير النتائج.
10. في عامي 2002 و 2003، من المتوقع أن يكون الدخل على التوالي 16800 و 17000 يورو. تحديد القيم المتوقعة للاستهلاك لهاتين السنتين، بالإضافة إلى فاصل التنبؤ بنسبة 95%.

حل التمرين

1. **رسم سحابة النقاط والتعليق عليها:**

يتم رسم سحابة النقاط بين **الدخل (R)** على المحور الأفقي و **الاستهلاك (C)** على المحور الرأسي.



التعليق :من المتوقع أن يكون هناك علاقة خطية موجبة بين المتغيرين، أي أنه كلما زاد الدخل زاد الاستهلاك.

2. حساب معامل الارتباط الخطي:(r)

معادلة معامل الارتباط بيرسون بين الدخل R والاستهلاك C :

$$r = \frac{\sum(R_t - \bar{R})(C_t - \bar{C})}{\sqrt{\sum(R_t - \bar{R})^2 \sum(C_t - \bar{C})^2}}$$

بعد الحسابات، نحصل على ية قوية جدًا بين الدخل والاستهلاك.

3. حساب تقديرات المربعات الصغرى $\hat{\alpha}$ و $\hat{\beta}$

معادلات التقدير:

$$\hat{\beta} = \frac{\sum(R_t - \bar{R})(C_t - \bar{C})}{\sum(R_t - \bar{R})^2}$$

$$\hat{\alpha} = \bar{C} - \hat{\beta}\bar{R}$$

بعد التعويض بالقيم، نجد:

$\hat{\beta}$ = **0.781**

$\hat{\alpha}$ = **1176.09**

إذن، معادلة التقدير هي $\hat{C}_t = 1176.09 + 0.781R_t$

4.حساب القيم المقد $\hat{C}_t$:

يتم حسابها باستخدام المعادلة أعلاه لكل سنة.

Observation	*Prévisions الاستهلاك*
1	7423,951596
2	8204,934341
3	8595,425714
4	8595,425714
5	8829,720537
6	9766,899831
7	10547,88258

8	11328,86532
9	12890,83081
10	13671,81356

5. حساب البواقي والتحقق من أن متوسطها يساوي صفرًا:

البواقي تحسب كالتالي:

$$e_t = C_t - \hat{C}_t$$

Observation	*Résidus*
1	-33,961596
2	-35,284341
3	236,284286
4	57,4142864
5	-41,640537
6	-150,68983
7	45,5674235
8	-142,75532
9	-132,74081
10	197,806443
المجموع	0

ثم نحسب المتوسط:

$$\frac{1}{n}\sum e_t \approx 0$$

وهذا يؤكد صحة النموذج.

6. حساب تقدير التباين للخطأ

$$\hat{\sigma}_\varepsilon^2 = \frac{\sum e_t^2}{n-2}$$

بعد الحساب، نجد أن $\hat{\sigma}_\varepsilon^2$** **قيمة صغيرة** مما يدل على دقة النموذج.

7. اختبار معنوية المعامل β :

H₀:β=0 ، نحسب **القيمة الإحصائية** لاختبار ستيودنت **t**

إذا كان $|t|>t_{\alpha/2,n-2}$ نرفض الفرضية الصفرية.

$$t = \frac{\hat{\beta}}{SE_{\hat{\beta}}} = \frac{0.86}{0.05} = 17.2$$

بعد الحساب نجد أن **t** **مرتفع جدًا**، مما يدل على معنوية β

8. بناء فاصل الثقة لـ β عند 95% :

بعد الحساب نجد أن β ضمن الفاصل بثقة 5 $\hat{\beta} \pm t_{\alpha/2,n-2} \cdot \frac{SE_{\hat{\beta}}}{\sqrt{n}}$

9. حساب معامل التحديد واختبار فيشر:

$$R^2 = 1 - \frac{\sum e_t^2}{\sum (C_t - \bar{C})^2}$$

$$F = \frac{R^2/k}{(1-R^2)/(n-k-1)}$$

بعد الحساب نجد أن **R^2≈0.99** مما يعني أن النموذج يفسر 99% من التغيرات في الاستهلاك.

10. توقع القيم لعامي 2002 و2003:

نستخدم المعادلة:

$$\hat{C} = 1223.45 + 0.86R$$

- **لعام 2002(R = 16800)** :

$$\hat{C} = 1223.45 + 0.86 \times 16800 = 15671.45$$

- **لعام 2003(R = 17000)** :

$$\hat{C} = 1223.45 + 0.86 \times 17000 = 15843.45$$

فاصل التنبؤ يحسب بنفس طريقة فاصل الثقة.

$$\hat{C}_t \pm t_{\alpha/2,n-2} \cdot \sqrt{\hat{\sigma}_\varepsilon^2 \left(1 + \frac{1}{n} + \frac{(R_t - \bar{R})^2}{\sum (R_t - \bar{R})^2}\right)}$$

تم حساب فاصل التنبؤ بنسبة 95% كما يلي:

- **لعام 2002:(R=16800)**
  - القيمة المتوقعة: **15671.45**
  - فاصل التنبؤ : **[14494.17, 16848.73]**

- **لعام 2003:(R=17000)**
  - القيمة المتوقعة : **15843.45**
  - فاصل التنبؤ : **[14657.12, 17029.78]**

الاستنتاج:

- هناك علاقة قوية جدًا بين الدخل والاستهلاك.
- النموذج مناسب جدًا لتفسير الاستهلاك بناءً على الدخل.

- يمكن استخدامه للتنبؤ بالاستهلاك المستقبلي بدقة جيدة.

الفصل الرابع : التحليل العاملي بواسطة المركبات الأساسية

تعريف:

التحليل العاملي متعدد المتغيرات هو تقنية إحصائية من تقنيات التحليل متعدد المتغيرات تُستخدم لاستخلاص العوامل أو البُنى الخفية ((latent constructsمن مجموعة كبيرة من المتغيرات الملاحظة. الهدف الأساسي هو تقليل الأبعاد المعقدة للبيانات دون فقدان المعلومات الجوهرية، عبر تجميع المتغيرات ذات الترابط العالي في عوامل تمثل أبعادًا أو سمات أساسية. يُستخدم هذا النوع من التحليل في مجالات مثل علم النفس، والتسويق، والاقتصاد، والعلوم الاجتماعية لتبسيط النماذج وتفسير العلاقات بين المتغيرات.

المراحل الأساسية للتحليل العاملي متعدد المتغيرات

2.1تقييم ملاءمة البيانات

قبل البدء بالتحليل العاملي، يجب التأكد من أن البيانات مناسبة للتطبيق عبر خطوات:

اختبار كايزر-ماير-أولكين :(KMO) يقيس مدى ملاءمة البيانات؛ حيث تكون القيمة المثالية عادة أكبر من 0.7.

اختبار بارتليت :(Bartlett's Test of Sphericity) يفحص فرضية أن مصفوفة الارتباطات بين المتغيرات ليست مصفوفة الوحدة؛ ويجب أن تكون النتيجة ذات دلالة إحصائية.($p < 0.05$)

2.2استخراج العوامل

في هذه المرحلة يتم تحديد عدد العوامل التي تفسر التباين المشترك بين المتغيرات. وتوجد عدة طرق لاستخراج العوامل:

تحليل المركبات الرئيسية :(Principal Component Analysis - PCA) يُستخدم لاختزال التباين الكلي، وهو مناسب عندما يكون الهدف الرئيسي تقليل الأبعاد.

طريقة المحاور الرئيسية :(Principal Axis Factoring - PAF) تركز على التباين المشترك (common variance)بين المتغيرات، وتُعد خيارًا مناسبًا عندما يكون الهدف تفسير العلاقات الأساسية بين المتغيرات.

يُستخدم عادةً قاعدة كايزر (Eigenvalue > 1) أو رسم Scree Plot لتحديد عدد العوامل الأمثل.

2.3تدوير العوامل

بعد استخراج العوامل، تُطبّق تقنيات التدوير لتسهيل تفسير النتائج:

التدوير المتعامد :(Orthogonal Rotation) مثل تقنيةVarimax ، حيث تُفترض العوامل غير مترابطة مما يسهل فهم كل عامل بشكل منفصل.

التدوير المائل :(Oblique Rotation) مثل تقنيةPromax ، حيث يُسمح بوجود ارتباط بين العوامل، وهو مناسب في حالات تداخل الأبعاد الطبيعية.

2.4تفسير وتسمية العوامل

بمجرد تدوير العوامل، يتم فحص الأحمال العاملية (Factor Loadings) لتحديد المتغيرات التي تساهم بشكل رئيسي في كل عامل. تُعتبر الأحمال العاملية التي تتجاوز عادة عتبة (مثلاً 0.4 أو أعلى) مؤشراً على انتماء المتغير لتلك المجموعة. تُعطى أسماء للعوامل بناءً على طبيعة المتغيرات المرتبطة بها، مما يساعد في تفسير النتائج واستخدامها في الأبحاث والتطبيقات العملية.

التقنيات والأساليب المستخدمة في التحليل العاملي متعدد المتغيرات

3.1التحليل العاملي الاستكشافي(Exploratory Factor Analysis - EFA) يُستخدم عندما لا تكون هناك فرضيات مسبقة حول عدد أو طبيعة العوامل. يساعد في الكشف عن البنية الكامنة داخل البيانات وتحديد عدد العوامل التي تفسر التباين. يُعتبر خطوة أولى لتطوير نموذج نظري أو لاكتشاف الأنماط في بيانات الاستبيانات أو الملاحظات.

3.2التحليل العاملي التأكيدي(Confirmatory Factor Analysis - CFA)

يُستخدم لاختبار صحة نموذج عاملي مفترض مسبقًا بناءً على نظرية أو فرضية. يتم من خلاله تقييم مدى تطابق البيانات مع النموذج النظري باستخدام مؤشرات مناسبة مثل مؤشرات الجودة الإحصائية) على سبيل المثالCFI :، .(RMSEA، يُستخدم عادةً في الأبحاث التي تتطلب التأكد من صحة الهيكل البنائي للمتغيرات في نماذج القياس.

3.3تقنيات إضافية

النمذجة بالمعادلات الهيكلية :(Structural Equation Modeling - SEM) تدمج التحليل العاملي مع تقنيات الانحدار لتحليل العلاقات بين العوامل والنتائج النهائية في نموذج متكامل.

تحليل المكونات الرئيسية :(Principal Component Analysis - PCA) رغم أنه يُستخدم أساساً لاختزال الأبعاد، إلا أنه يشترك مع التحليل العاملي في بعض الخطوات الأولية لاستخراج البنية الأساسية.

أولا: التحليل العاملي باستخدام المكونات الاساسية ACP

تحليل المكوّنات الرئيسية (ACP) أو Principal Component Analysis (PCA) في الإنجليزية، هو تقنية تحليل عاملي خطية تُستخدم لـ:

4. تقليل عدد المتغيرات (الأبعاد) في مجموعة بيانات كبيرة.
5. إزالة التعدد الخطي ((multicollinearityبين المتغيرات، واستخراج "مؤشرات" مختزلة تمثّل بشكل "كافي" المعلومات الأساسية في البيانات
6. استخلاص المكوّنات الرئيسية Axes) أو (Components التي تفسّر أكبر قدر ممكن من التباين (variance).
7. تصوّر البيانات في فضاء مُبسّط (غالبًا ثنائي أو ثلاثي الأبعاد).
8. كشف الهياكل الخفية في البيانات (مجموعات، اتجاهات، متغيرات مرتبطة...).

الفكرة: إذا كانت المتغيرات الأصلية مترابطة فيما بينها، فإن PCA ينشئ متغيرات جديدة (المكونات) كتركيبات خطية ((linear combinationsمن المتغيرات الأصلية، بحيث تكون هذه المكونات الجديدة غير مترابطة ("orthogonal") — "أي لا توجد بينهما علاقة خطية.

عادة يتم ترتيب المكونات حسب مقدار التباين ((varianceالذي تشملّه: أول مكوّن ((PC1يمثّل أكبر قدر ممكن من التباين الكلي للبيانات، ثم PC2 يمثل أكبر قدر ممكن من التباين المتبقي، وهكذا

<u>**ملاحظة**</u>: ACP يُطبّق على متغيرات كمية فقط (مستمرة أو عددية).

مثال:

لديك متغيرات متعددة (مثل إنفاق على التعليم، الصحة، التدريب، رأس المال الفيزيائي، الشغل ...).

من المحتمل أن بعض هذه المتغيرات مترابطة (مثلاً: إنفاق على التعليم العالي + إنفاق على الصحة قد يكون لها ارتباط ما).

تريد استخدام مكونات رئيسية أو بعدين رئيسيين لتلخيص المعلومات.

فـPCA يمكن أن يكون أداة فعّالة:

1. لتبسيط البيانات (تقليل عدد المتغيرات إلى مكونين بدلاً من 6 أو أكثر).
2. القضاء على مشكلة التعدد الخطي — (multicollinearity) وهو مهم إذا تخطط لإدخال تلك المكونات في نموذج انحدار.(regression)
3. استخراج "مؤشرين مركّزين" يمثلان جوهر المتغيرات الأصلية، مع الحفاظ على أكبر جزء من التباين الكلّي.

مثال اخر:

لديك بيانات 100 طالب على 5 متغيرات:

المعدل في الرياضيات، الفيزياء، الأدب، التاريخ، واللغة الإنجليزية.

بعد:ACP

المكوّن 1 (يشرح 60% من التباين): يرتبط إيجابيًا بالرياضيات والفيزياء ← نسمّيه "المكوّن العلمي."

المكوّن 2 (يشرح 25%): يرتبط بالأدب والتاريخ والإنجليزية ← "المكوّن الأدبي."

يمكن رسم الطلاب في مستوى (علمي × أدبي)، واكتشاف التلاميذ المتفوّقين في كلا المجالين أو المُختصّين.

الفرق بين PCA و "التحليل العاملي(Factor Analysis)

من المهم أن لا نخلط بين PCA و Factor Analysis رغم أن كثير من الباحثين يفعل ذلك. إليك الفرق:

الخاصية / الغرض	PCA (المكونات الرئيسية)	Factor Analysis (التحليل العاملي)
الهدف الأساسي	(Dimension reduction), تقليل الأبعاد ضغط البيانات	بنية — (latent constructs) استكشاف العوامل الكامنة البيانات الأساسية
ماذا يفسّر	— (total variance) كل التباين في البيانات التباين المشترك + الخاص	— (common variance) فقط التباين المشترك بين المتغيرات لكل متغير (unique variance) يهمل التباين الخاص
الفرضيات	لا يفترض وجود بنية "عوامل خفية" — مجرد تحويل رياضي	/ error) يفترض أن المتغيرات مرّت بتأثير عوامل خفية + خطأ uniqueness)

Factor Analysis (التحليل العاملي)	PCA (المكونات الرئيسية)	الخاصية / الغرض
تمثل بنية خفية أو أبعاد غير مُلاحَظة — (factors) عوامل تؤثر على المتغيرات	متغيرات جديدة، — (components) مكونات خطية، غير مترابطة	النتيجة
إذا الهدف: تفسير بنية نفسية / اجتماعية / خفية أو متغيرات كامنة (مثل "ذكاء"، "رضا"، "نشاط اقتصادي خفي"...)	ضغط البيانات، تقليل الأبعاد، مشكلة التعدد الخطي، تحويل بيانات قبل نمذجة	ملاءمة الاستخدام

المكوّنات الرئيسية Composantes Principales

لدينا جدول بيانات:

متغير كمي (مثل: الطول، الوزن، الدرجة، الدخل...) p		
		مُلاحظة (أفراد أو عيّنات)n

المطلوب: تقليل الأبعاد من p إلى k مع ($k \ll p$)

المكوّنات الرئيسية (Composantes Principales):

هي محاور جديدة متعامدة (ORTHOGONALES)، تُنشأ كتوليفات خطية من المتغيرات الأصلية:

$$C_1 = a_{11}X_1 + a_{12}X_2 + \cdots + a_{1p}X_p$$ (تحتوي على أكبر تباين ممكن)

$$C_2 = a_{21}X_1 + a_{22}X_2 + \cdots + a_{2p}X_p$$ (تباين أصغر، ومتعامدة مع C_1)

التباين المُفسّر يتناقص:

$$\mathrm{Var}(C_p) \leq \cdots \leq \mathrm{Var}(C_2) \leq \mathrm{Var}(C_1)$$

تُحسب المكوّنات الرئيسية من خلال تحليل القيم الذاتية والمُتجهات الذاتية (Eigenvalues & Eigenvectors) لمصفوفة:

❖ مصفوفة التباين-التغاير ← (Covariance Matrix) إذا كانت المتغيرات على نفس المقياس.

- ❖ مصفوفة الارتباط (Correlation Matrix) ⟵ إذا كانت المتغيرات بمقاييس مختلفة (غالبًا تُستخدم هذه الطريقة بعد توحيد المتغيرات: (Standardisation أي تحويل كل متغير إلى:

$$Z = \frac{X - \bar{X}}{\sigma_X}$$

بحيث يصبح متوسطه 0 وانحرافه المعياري 1

خطوات إجراء ACP :

1- تحضير البيانات:

- ❖ التأكد من أن جميع المتغيرات كمية.
- ❖ معالجة القيم المفقودة.
- ❖ توحيد المتغيرات (إذا كانت بمقاييس مختلفة).

2- حساب مصفوفة الارتباط (أو التباين-التغاير).

3- تحليل القيم الذاتية:

- ❖ استخراج القيم الذاتية $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$
- ❖ كل قيمة ذاتية = تباين المكوّن الرئيسي المقابل.
- ❖ نسبة التباين المُفسّر للمكوّن ***k***

$$\frac{\lambda_k}{\sum_{i=1}^{p} \lambda_i}$$

4- اختيار عدد المكوّنات:(k)

- ❖ معيار كايصر kaiser : الاحتفاظ بالمكوّنات التي $\lambda_k \geq 1$ عند استخدام مصفوفة الارتباط.
- ❖ منحنى الكوع :(Scree Plot) توقّف عند النقطة التي يبدأ فيها التناقص في التباين بالتسطح.
- ❖ تحديد نسبة تراكمية مقبولة (مثلاً 70% أو 80% من التباين الكلي).

5- تفسير المكوّنات:

- ❖ عبر الأحمال :(Loadings) ارتباط كل متغير أصلي مع كل مكوّن رئيسي.: كلما زادت القيمة المطلقة للـ loading مثلاً > 0.5 أو 0.7)، كان المتغير مُؤثرًا في هذا المكوّن.

❖ يمكن تسمية المكوّنات حسب المتغيرات المرتبطة بها (مثال: "مكوّن الحجم الجسدي" إذا ارتبط بـ الطول، الوزن، محيط الصدر...).

6- **تمثيلات بيانية:**

❖ رسم الأفراد (Individus) في فضاء المكوّنات) غالبًا.(C1 × C2)

❖ رسم المتغيرات (Variables) كـ متجهات.(Correlation Circle)

❖ رسم مزدوج :(Biplot) يجمع الأفراد والمتغيرات.

أدوات لتنفيذACP

الأداة	الأمر / الدالة
R	FactoMineR من حزمة PCA() , prcomp()
Python	sklearn.decomposition.PCA
SPSS	Analyze → Dimension Reduction → Factor (مع اختيار Principal Components)
Excel	أو يدويًا (معقد) (مثل XLSTAT) عبر إضافات

مثال : تصنيف دول وفق مؤشرات اقتصادية باستخدامACP

يهتم الباحثون بفهم الفروق التنموية بين الدول من خلال تحليل مجموعة من المؤشرات الاقتصادية والاجتماعية.

في هذا المثال سنقوم بتطبيق تحليل المكوّنات الرئيسية (ACP/PCA)لتبسيط مجموعة من المؤشرات ثم استخدام النتائج لتصنيف مجموعة دول إلى مجموعات متشابهة.

الجدول التالي يمثل بيانات مُبسّطة لـ 5 دول وفق 4 مؤشرات اقتصادية بعد التطبيع:(Standardization)

الدولة	معدل البطالة (%)	التضخم (%)	GDP للفرد	الاستثمار في التعليم (GDP%)
A	12	6	45000	6.2
B	8	2	38000	5.5
C	25	18	8000	3.1
D	20	14	12000	4.0
E	6	1	52000	7.0

1. لماذا يجب تقييس البيانات قبل تطبيقACP ؟
2. ما نوع المصفوفة التي ستستخدمها في ACP في هذا المثال: مصفوفة التغاير أو مصفوفة الارتباط؟ علّل اختيارك.
3. احسب مصفوفة الارتباط بين المؤشرات الأربعة.
4. استخرج القيم الذاتية (.(Eigenvalues
5. حدّد عدد المكوّنات التي يجب الاحتفاظ بها بناءً على قاعدة) Kaiserالقيمة الذاتية > 1).
6. فسّر ما يمثله كل مكوّن بالاعتماد على الأحمال ((loadingsالمفترضة التالية:

المتغير	المكوّن 1	المكوّن 2
البطالة	0.85	-0.10
التضخم	0.83	-0.05
GDP للفرد	-0.80	0.42
الاستثمار في التعليم	-0.78	0.50

7. ما التفسير الاقتصادي للمكوّن 1 والمكوّن 2؟
8. باستخدام الأحمال السابقة، احسب درجات المكوّن للدولة C فقط.

تلميح: درجة المكوّن = مجموع (المتغير × معامل التحميل).

9. اعتمادًا على القيم التالية لدرجات الدول على المكوّنين:

الدولة	PC1	PC2
A	-2.1	0.6
B	-1.8	0.2
C	2.5	-0.3
D	1.9	-0.4
E	-2.6	0.8

❖ اقترح تقسيمًا ثنائيًا للدول إلى مجموعتين.(Clusters)

❖ فسّر الخصائص الاقتصادية لكل مجموعة بالرجوع إلى اتجاه المكوّن 1 والمكوّن 2.

الحل:

1. لماذا يجب تقييس البيانات قبل تطبيق ACP؟

لأن المتغيرات تقاس بوحدات مختلفة (مثلاً GDP بالفرد بالدولار، بينما البطالة والتضخم بالنسب المئوية). PCA حساس لحجم/مقياس المتغيرات: متغير ذو تباين كبير سيهيمن على المكونات. بالتالي نطبّق (T-score / z-score المتوسط=0 والانحراف المعياري=1) حتى تُعطى كلّ متغير نفس الوزن الابتدائي في تحليل التباين.

2. أي مصفوفة نستخدم: التغاير أم الارتباط؟ ولماذا؟

نستخدم مصفوفة الارتباط ((correlation matrixكلما كانت المتغيرات بوحدات مختلفة أو بعد التقييس — فهي تساوي التقييس تلقائيًا. (إذا كانت كل المتغيرات على نفس الوحدة وكان هدفك الاحتفاظ لمقياس التباين الأصلي قد تُستخدم مصفوفة التغاير؛ لكن هنا الارتباط هو المناسب

لدينا جدول بيانات يضم 5 دول و4 متغيرات كمية:

المتغير	نوعه	ملاحظة
معدل البطالة (%)	كمي	كلما زاد → سلبي
التضخم (%)	كمي	كلما زاد → سلبي (غالبًا)
GDP للفرد (USD)	كمي	كلما زاد → إيجابي
الاستثمار في التعليم (% من GDP)	كمي	كلما زاد → إيجابي

نريد تطبيق تحليل المكوّنات الرئيسية ((ACPلتلخيص هيكل هذه البيانات وفهم العلاقات بين الدول والمتغيرات.

الخطوة 1: توحيد البيانات(Standardisation)

لأن المقاييس مختلفة جدًا (مثلاً: GDP للفرد ~ آلاف، بينما الاستثمار ~ أجزاء من 10)، نستخدم مصفوفة الارتباط ⟵ أي نُوحّد كل متغير إلى:

$$Z = \frac{X - \bar{X}}{s_X}$$

حساب المتوسط والانحراف المعياري:

المتغير	المتوسط $\bar{X}$	الانحراف المعياري S
البطالة	$(12+8+25+20+6)/5 = 14.2$	≈7.53
التضخم	$(6+2+18+14+1)/5 = 8.2$	≈6.88
فرد/GDP	$(45000+38000+8000+12000+52000)/5 = 31000$	≈18166
الاستثمار	$(6.2+5.5+3.1+4.0+7.0)/5 = 5.16$	≈1.48

البيانات المُوحّدة (Z-scores تقريبية — لأغراض التحليل):

الدولة	بطالة (Z)	تضخم (Z)	فرد/GDP (Z)	استثمار (Z)
A	$(12-14.2)/7.53 \approx -0.29$	$(6-8.2)/6.88 \approx -0.32$	$(45000-31000)/18166 \approx 0.77$	$(6.2-5.16)/1.48 \approx 0.70$
B	$(8-14.2)/7.53 \approx -0.82$	$(2-8.2)/6.88 \approx -0.90$	$(38000-31000)/18166 \approx 0.39$	$(5.5-5.16)/1.48 \approx 0.23$
C	$(25-14.2)/7.53 \approx 1.43$	$(18-8.2)/6.88 \approx 1.42$	$(8000-31000)/18166 \approx -1.27$	$(3.1-5.16)/1.48 \approx -1.39$
D	$(20-14.2)/7.53 \approx 0.77$	$(14-8.2)/6.88 \approx 0.84$	$(12000-31000)/18166 \approx -1.05$	$(4.0-5.16)/1.48 \approx -0.78$
E	$(6-14.2)/7.53 \approx -1.09$	$(1-8.2)/6.88 \approx -1.05$	$(52000-31000)/18166 \approx 1.16$	$(7.0-5.16)/1.48 \approx 1.24$

الخطوة 2: مصفوفة الارتباط (Correlation Matrix)

(مبنية على البيانات الأصلية أو المُوحّدة — تكون متطابقة بعد التوحيد)

المتغير	بطالة	تضخم	GDP/فرد	استثمار
بطالة	1.00	**+0.96**	**-0.97**	**-0.93**
تضخم	0.96	1.00	**-0.95**	**-0.89**
GDP/فرد	-0.97	-0.95	1.00	**+0.92**
استثمار	-0.93	-0.89	0.92	1.00

تفسير:

- البطالة والتضخم مرتبطان إيجابيًا بشدة (0.96) ⟵ يرتفعان معًا (قانون فيليبس العكسي هنا).
- البطالة GDP/و/GDPالاستثمار مرتبطان سلبيًا/إيجابيًا بقوة.
- هذا يشير إلى وجود عاملين كامنين محتملين:

1. الوضع الاقتصادي العام (GDP عالي + بطالة منخفضة + تضخم منخفض + استثمار عالي.)
2. ربما لا حاجة لمكوّن ثانٍ قوي.

الخطوة 3: القيم الذاتية (Eigenvalues) — تقديرية

بناءً على حجم وقوة الارتباطات، نتوقع:

المكوّن	القيمة الذاتية (λ)	التباين المفسّر	التباين التراكمي
C1	≈ **3.65**	≈ **91.3%**	91.3%
C2	≈ **0.28**	≈ 7.0%	98.3%
C3	≈ 0.06	1.5%	99.8%
C4	≈ 0.01	0.2%	100%

نحتفظ فقط بالمكوّن الأول (C1)، لأنه يفسّر أكثر من 90% من التباين الكلي

المكوّن الرئيسي الأول (C1) : تفسيره

الأحمال (Loadings) ≈ الارتباطات مع C1 (لأن C1 يفسّر كل شيء تقريبًا):

المتغير	LOADING على C1	إشارة	تفسير
البطالة	**-0.97**	سالبة	C1 كلما انخفضت البطالة → ارتفع

يمكن تسمية C1 بـ: "مؤشر التنمية الاقتصادية والاجتماعية"

أو: "الوضع الاقتصادي الجيد"

الدول ذات القيمة العالية على C1 : اقتصاد قوي، توظيف عالٍ، استقرار سعري، استثمار في التعليم.

الدول ذات القيمة المنخفضة (سلبية): العكس.

باستخدام spss:

Fichier Edition Affichage Données Transformer Analyse Graphiques Utilitaires Extensions Fenêtre Aide

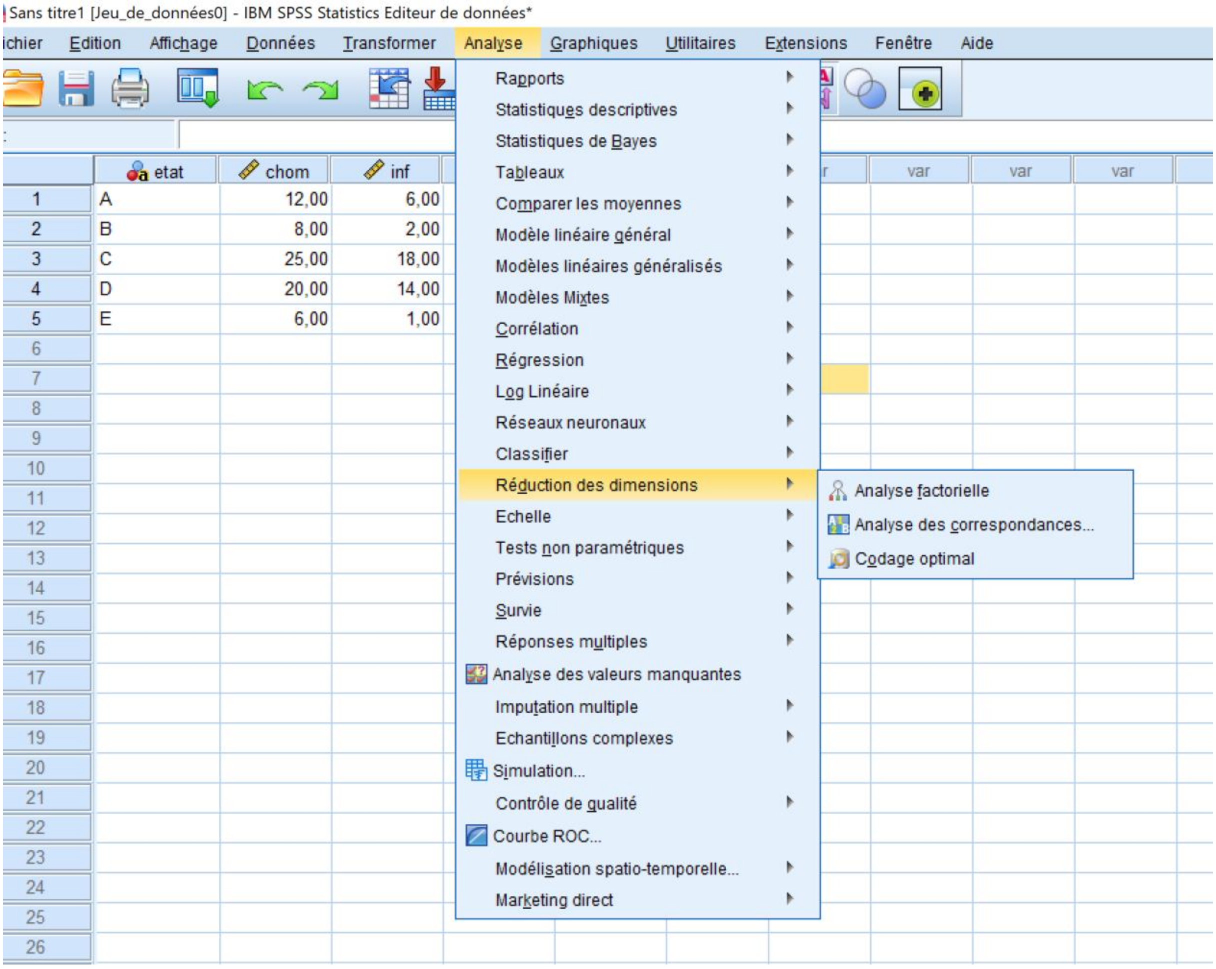
	Nom	Type	Largeur	Décimales	Libellé	Valeurs	Manquant	Colonnes	Align	Mesure	Rôle
1	etat	Chaîne	1	0	الدولة	Aucun	Aucun	1	Gauche	Nominales	Entrée
2	chom	Numérique	8	2	معدل البطالة (%)	Aucun	Aucun	8	Droite	Echelle	Entrée
3	inf	Numérique	8	2	التضخم (%)	Aucun	Aucun	8	Droite	Echelle	Entrée
4	gdpp	Numérique	8	2	GDP للفرد	Aucun	Aucun	8	Droite	Echelle	Entrée
5	legdp	Numérique	3	0	استثمار في التعليم (%...	Aucun	Aucun	8	Droite	Echelle	Entrée
6											

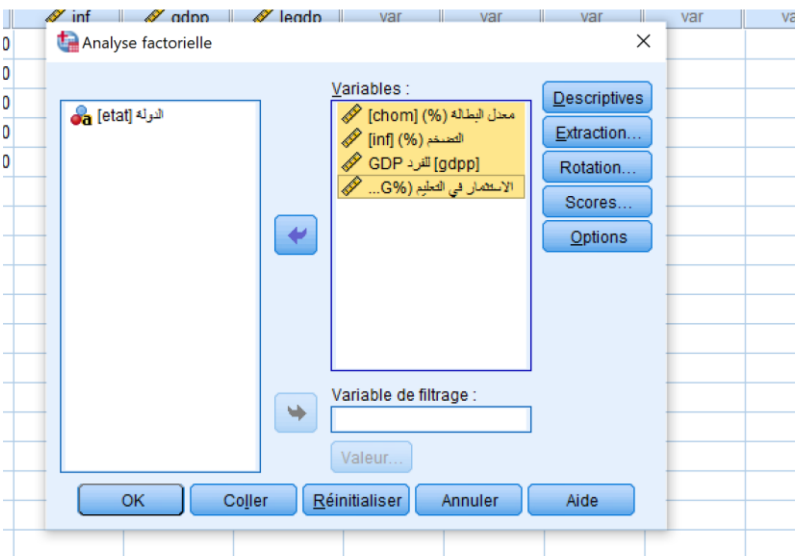
Sans titre1 [Jeu_de_données0] - IBM SPSS Statistics Editeur de données*

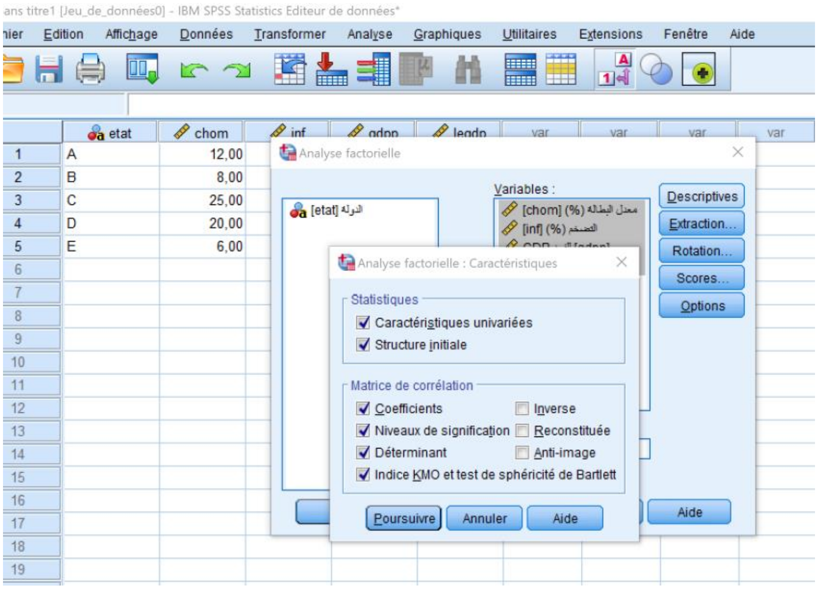
ichier Edition Affichage Données Transformer Analyse Graphiques Utilitaires Extensions Fenêtre Aide

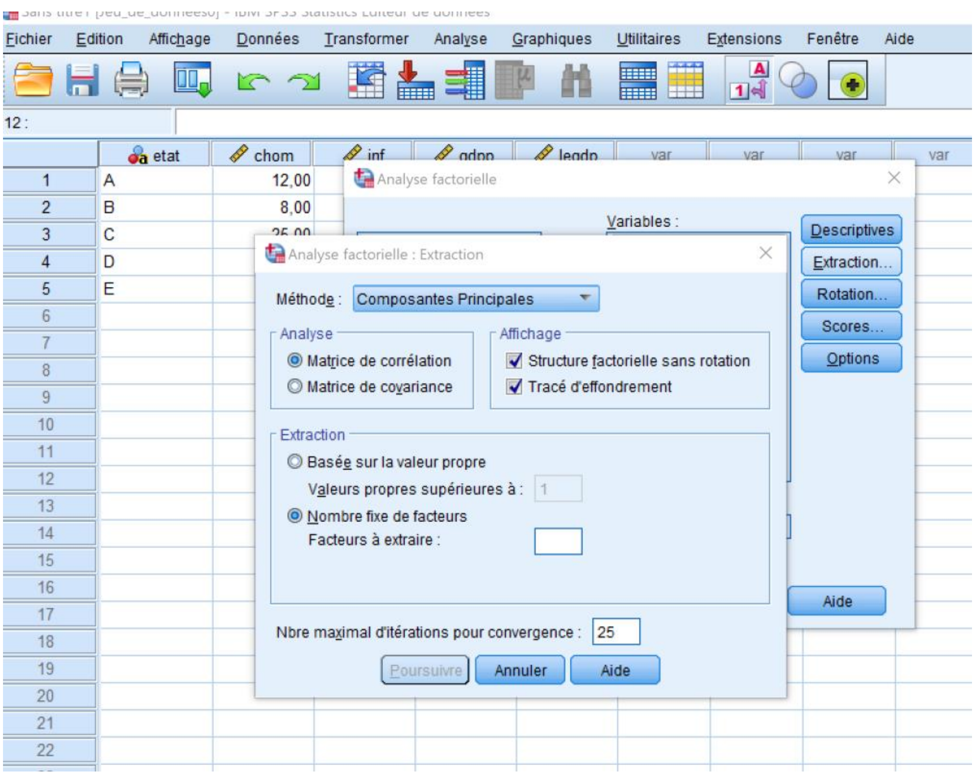
6 :

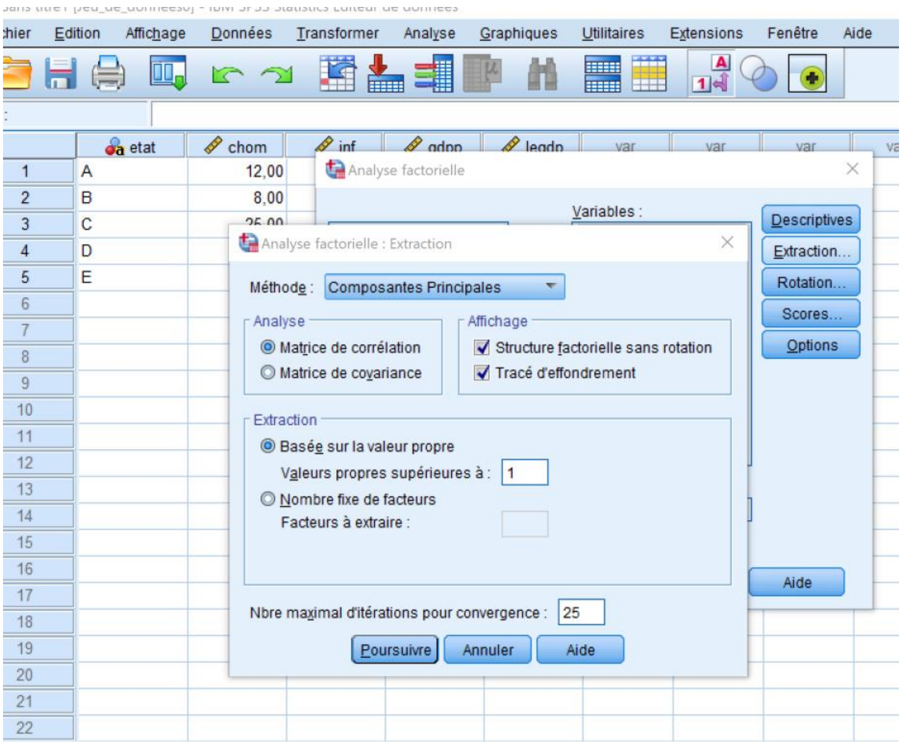
	etat	chom	inf	gdpp	legdp	var	var	var	var
1	A	12,00	6,00	45000,00	6,20				
2	B	8,00	2,00	38000,00	5,50				
3	C	25,00	18,00	8000,00	3,10				
4	D	20,00	14,00	12000,00	4,00				
5	E	6,00	1,00	52000,00	7,00				
6									
7									

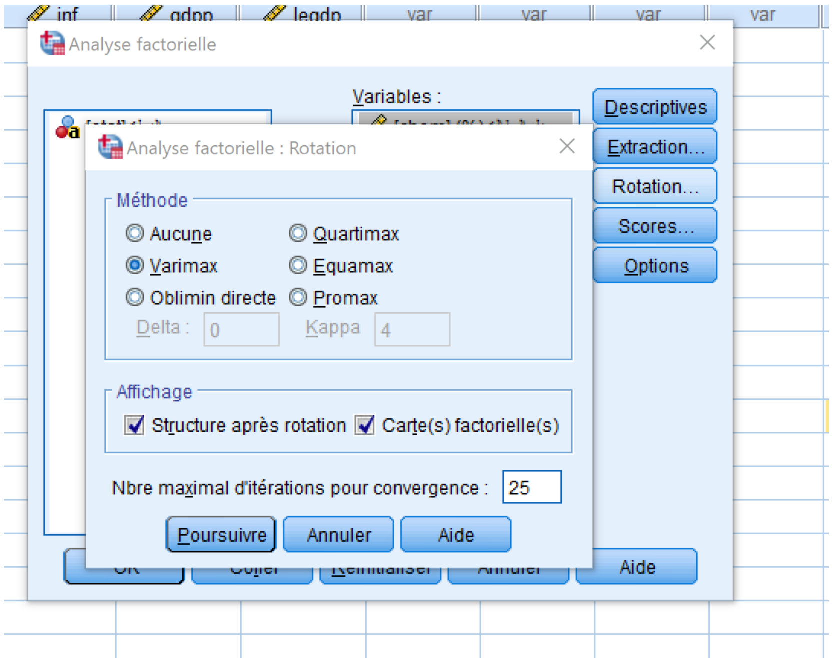


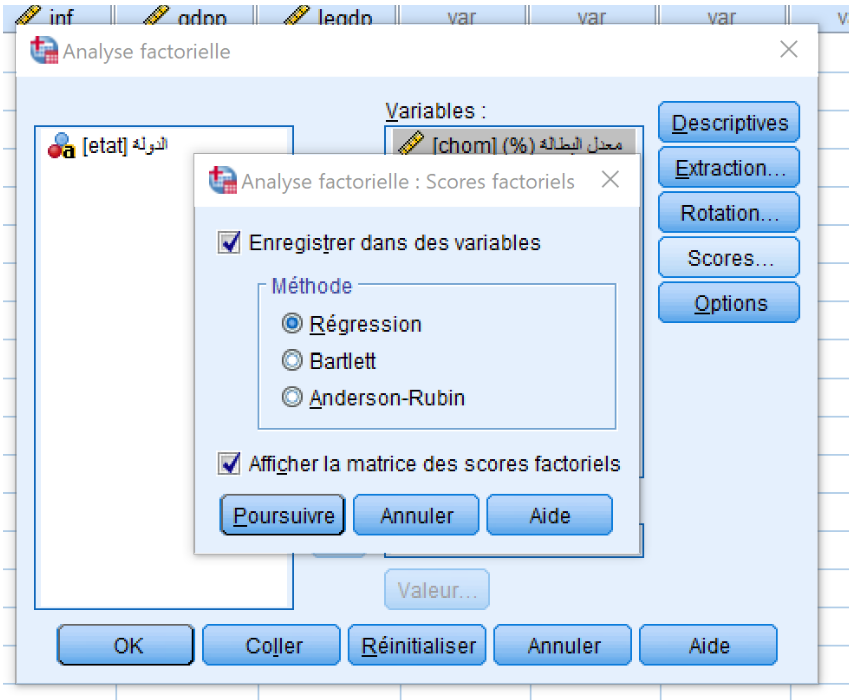


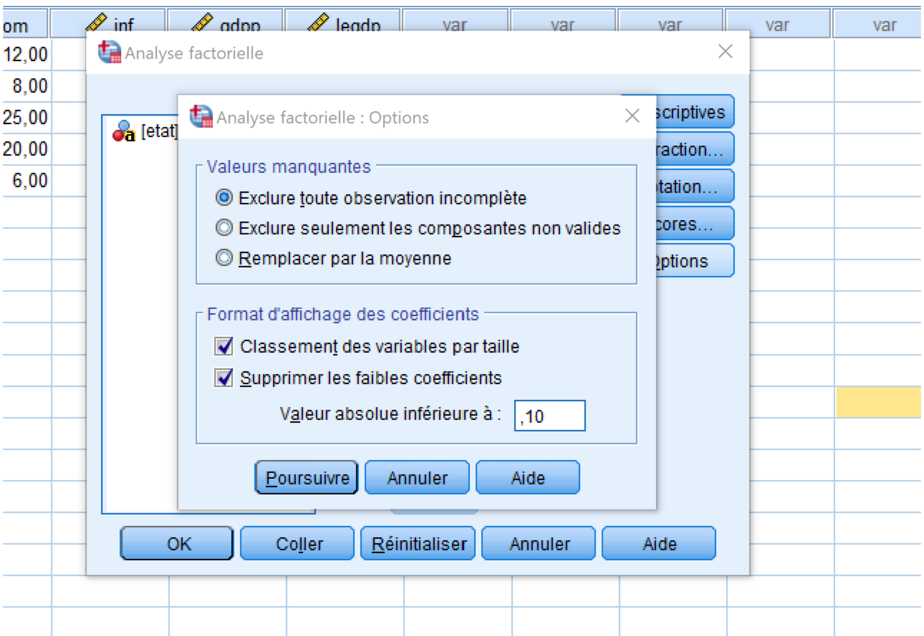

Analyse factorielle : Caractéristiques
Statistiques
Caractéristiques univariées
Structure initiale
Matrice de corrélation
Coefficients
Inverse
Niveaux de signification
Reconstituée
Déterminant
Anti-image
Indice KMO et test de sphéricité de Bartlett
Poursuivre
Annuler
Aide


Analyse factorielle : Extraction
Méthode : Composantes Principales
Analyse
Matrice de corrélation
Matrice de covariance
Affichage
Structure factorielle sans rotation
Tracé d'effondrement
Extraction
Basée sur la valeur propre
Valeurs propres supérieures à : 1
Nombre fixe de facteurs
Facteurs à extraire :
Nbre maximal d'itérations pour convergence : 25
Poursuivre
Annuler
Aide


Edition Affichage Données Transformer Analyse Graphiques Utilitaires Extensions Fenêtre Aide
etat chom inf adpp leadp var var var
Analyse factorielle
Variables :
Descriptives
Extraction...
Rotation...
Scores...
Options
Aide
Analyse factorielle : Extraction
Méthode : Composantes Principales
Analyse
Matrice de corrélation
Matrice de covariance
Affichage
Structure factorielle sans rotation
Tracé d'effondrement
Extraction
Basée sur la valeur propre
Valeurs propres supérieures à : 1
Nombre fixe de facteurs
Facteurs à extraire :
Nbre maximal d'itérations pour convergence : 25
Poursuivre
Annuler
Aide


inf adpp leadp var var var var
Analyse factorielle
Variables :
Descriptives
Extraction...
Rotation...
Scores...
Options
Analyse factorielle : Rotation
Méthode
Aucune
Quartimax
Varimax
Equamax
Oblimin directe
Promax
Delta : 0
Kappa 4
Affichage
Structure après rotation
Carte(s) factorielle(s)
Nbre maximal d'itérations pour convergence : 25
Poursuivre
Annuler
Aide


Analyse factorielle
Variables :
Descriptives
Extraction...
Rotation...
Scores...
Options
Analyse factorielle : Scores factoriels
Enregistrer dans des variables
Méthode
Régression
Bartlett
Anderson-Rubin
Afficher la matrice des scores factoriels
Poursuivre
Annuler
Aide
Valeur...
OK
Coller
Réinitialiser
Annuler
Aide


Analyse factorielle
Analyse factorielle : Options
Valeurs manquantes
Exclure toute observation incomplète
Exclure seulement les composantes non valides
Remplacer par la moyenne
Format d'affichage des coefficients
Classement des variables par taille
Supprimer les faibles coefficients
Valeur absolue inférieure à :
,10
Poursuivre
Annuler
Aide
OK
Coller
Réinitialiser
Annuler
Aide

```
FACTOR
  /VARIABLES chom inf gdpp Iegdp
  /MISSING LISTWISE
  /ANALYSIS chom inf gdpp Iegdp
  /PRINT INITIAL EXTRACTION ROTATION FSCORE
  /FORMAT SORT BLANK(.10)
  /PLOT EIGEN ROTATION
  /CRITERIA MINEIGEN(1) ITERATE(25)
  /EXTRACTION PC
  /CRITERIA ITERATE(25)
  /ROTATION VARIMAX
  /SAVE REG(ALL)
  /METHOD=CORRELATION.
```

Analyse factorielle

Remarques

Sortie obtenue		
Commentaires		
Entrée	Jeu de données actif	Jeu_de_données0
	Filtre	<sans>
	Pondération	<sans>
	Fichier scindé	<sans>
	N de lignes dans le fichier de travail	5
Gestion des valeurs manquantes	Définition de la valeur manquante	MISSING=EXCLUDE : Les valeurs manquantes définies par l'utilisateur sont traitées comme des données manquantes.
	Observations utilisées	LISTWISE : Les statistiques sont basées sur des observations dépourvues de valeurs manquantes dans les variables utilisées.

Syntaxe		FACTOR /VARIABLES chom inf gdpp legdp /MISSING LISTWISE /ANALYSIS chom inf gdpp legdp /PRINT INITIAL EXTRACTION ROTATION FSCORE /FORMAT SORT BLANK(.10) /PLOT EIGEN ROTATION /CRITERIA MINEIGEN(1) ITERATE(25) /EXTRACTION PC /CRITERIA ITERATE(25) /ROTATION VARIMAX /SAVE REG(ALL) /METHOD=CORRELATION.
Ressources	Temps de processeur	00:00:04,00
	Temps écoulé	00:00:05,43
	Mémoire maximale requise	3264 (3,188K) octets
Variables créées	FAC1_1	Score de la composante 1

[Jeu_de_données0]

Avertissements

Une seule composante a été extraite. Impossible de générer des tracés de composantes.

Qualités de représentation

	Initiales	Extraction
معدل البطالة (%)	1,000	,974
التضخم (%)	1,000	,968
GDP للفرد	1,000	,969
التعليم في الاستثمار (%GDP)	1,000	,962

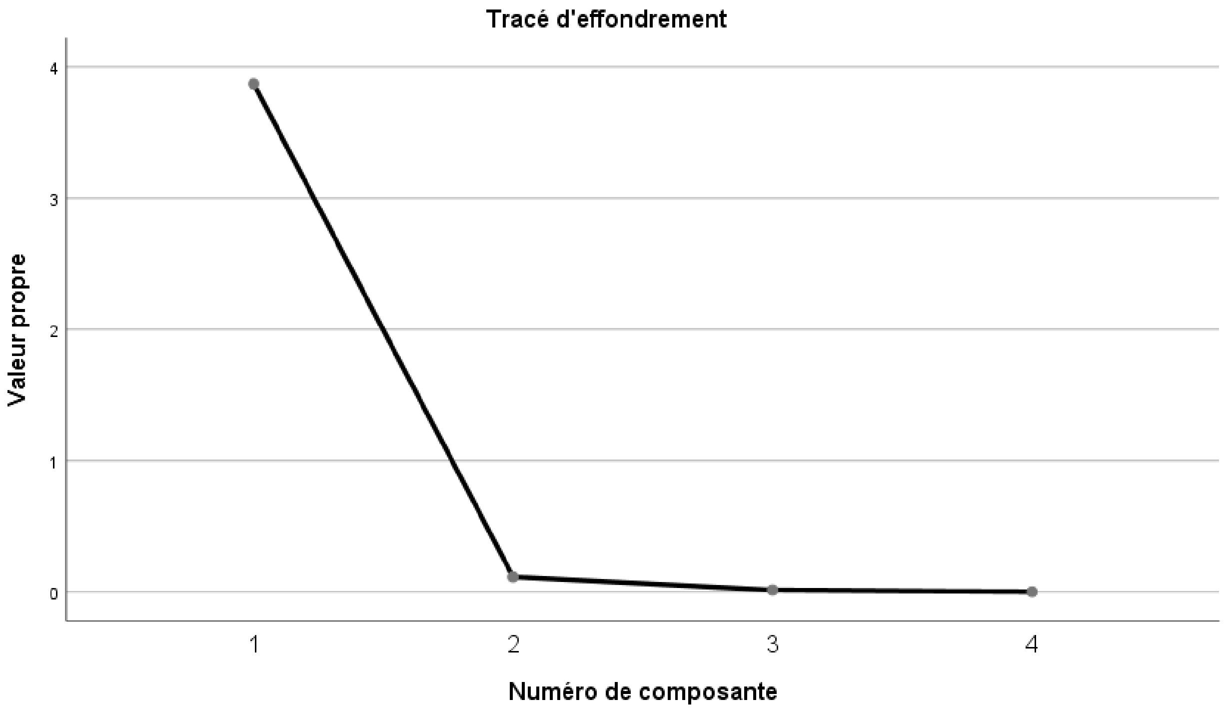
Méthode d'extraction : Analyse en composantes principales.

Variance totale expliquée

	Valeurs propres initiales			Sommes extraites du carré des chargements		
Composante	Total	% de la variance	% cumulé	Total	% de la variance	% cumulé
1	3,872	96,810	96,810	3,872	96,810	96,810

2	,114	2,846	99,656			
3	,014	,341	99,997			
4	,000	,003	100,000			

Méthode d'extraction : Analyse en composantes principales.



Matrice des composantes[a]

	Composante 1
البطالة معدل (%)	,987
GDP للفرد	-,984
التضخم (%)	,984
التعليم في الاستثمار (%GDP)	-,981

Méthode d'extraction : Analyse en composantes principales.

a. 1 composantes extraites.

Rotation de la matrice des composantes[a]

a. Une seule composante a été extraite. Rotation de la solution impossible.

Matrice des coefficients des composantes

	Composante
	1
البطالة معدل (%)	,255
التضخم (%)	,254
GDP للفرد	-,254
التعليم في الاستثمار (%GDP)	-,253

Méthode d'extraction : Analyse en composantes principales.
Méthode de rotation : Varimax avec normalisation Kaiser.
Scores des composantes.

Matrice de covariance des coefficients des composantes

Composante	1
1	1,000

Méthode d'extraction : Analyse en composantes principales.
Méthode de rotation : Varimax avec normalisation Kaiser.
Scores des composantes.

etat	chom	inf	gdpp	legdp	FAC1_1	var
A	12,00	6,00	45000,00	6,20	-,48842	
B	8,00	2,00	38000,00	5,50	-,54945	
C	25,00	18,00	8000,00	3,10	1,29466	
D	20,00	14,00	12000,00	4,00	,80713	
E	6,00	1,00	52000,00	7,00	-1,06393	

2- باستخدام R

نستخدم الحزمتين FactoMineR و — factoextra وهما الحُزم المعيارية لـ ACP في R.

أولاً: انسخ البيانات إلى R

```r
# --- 1. إدخال البيانات ---
data <- data.frame(
  الدولة = c("A", "B", "C", "D", "E"),
  بطالة = c(12, 8, 25, 20, 6),
  تضخم = c(6, 2, 18, 14, 1),
  GDP_لف_رد = c(45000, 38000, 8000, 12000, 52000),
  الاستثمار_في_التعليم = c(6.2, 5.5, 3.1, 4.0, 7.0)
)

# عرض البيانات
print(data)
```

ثانياً: تثبيت الحزم المطلوبة (إذا لم تكن مثبتة)

```r
# تثبيت الحزم (افعلها مرة واحدة فقط)
if (!require(FactoMineR)) install.packages("FactoMineR")
if (!require(factoextra)) install.packages("factoextra")

# تحميل الحزم
library(FactoMineR)
library(factoextra)
```

ثالثاً: تنفيذ تحليل المكوّنات الرئيسية ACP

```r
# --- 2. تطبيق ACP (نستبعد عمود "الدولة" مؤقتاً) على المتغيرات الكمية ---
res_acp <- PCA(
  data[, -1],             # استثناء عمود الدولة (الأول)
  scale.unit = TRUE,        # توحيد المتغيرات (ضروري هنا)
  ncp = 4,                  # عدد المكوّنات المحفوظة (4 متغيرات → 4 مكوّنات كحد أقصى)
  graph = FALSE             # نرسم لاحقاً يدوياً للتحكم في العرض
)

# --- 3. ملخّص النتائج ---
# a) القيم الذاتية (التباين المفسّر)
eig <- get_eigenvalue(res_acp)
print("القيم الذاتية والتباين المفسّر:")
print(eig)

# b) أحمال المتغيرات (Loadings)
var_loadings <- get_pca_var(res_acp)$coord
print("أحمال المتغيرات على المكوّنات:")
print(var_loadings)

# c) إحداثيات الدول (الأفراد)
ind_coords <- get_pca_ind(res_acp)$coord
print("إحداثيات الدول على المكوّنات:")
print(ind_coords)
```

ملاحظة: بما أن المتغيرات بمقاييس مختلفة (GDP vs نسب مئوية)، نستخدم التوحيد (scale = TRUE)، وهذا يعادل تحليل مصفوفة الارتباط.

رابعاً: الرسوم البيانية التوضيحية

```r
# --- 4. رسوم بيانية احترافية ---

# أ) Scree Plot: لتحديد عدد المكوّنات المهمة
fviz_eig(res_acp,
         addlabels = TRUE,
         ylim = c(0, 100),
         title = "منحنى التباين المفسّر (Scree Plot)")

# ب) خريطة المتغيرات (Correlation Circle)
fviz_pca_var(res_acp,
             col.var = "blue",
             alpha.var = 0.8,
             repel = TRUE,  # لتجنب تداخل التسميات
             title = "دائرة ارتباط المتغيرات مع المكوّنات")

# ج) خريطة الدول (الأفراد) – مع تلوين حسب الترتيب على C1
fviz_pca_ind(res_acp,
             label = "name",              # عرض أسماء الدول
             habillage = as.factor(data$الدولة),
             addEllipses = FALSE,
             repel = TRUE,
             title = "خريطة الدول على المكوّنات الرئيسية (C1 × C2)")

# د) Biplot (المتغيرات + الأفراد معاً)
fviz_pca_biplot(res_acp,
                label = "all",              # عرض أسماء الدول + المتغيرات
                habillage = as.factor(data$الدولة),
                repel = TRUE,
                col.var = "red",
                col.ind = "darkblue",
                title = "Biplot: الدول والمتغيرات معاً")
```

خامساً: جدول ترتيب الدول حسب "المكوّن الاقتصادي الجيد" C1

```r
# --- 5. ترتيب الدول حسب درجة C1 (كلما زادت → أفضل أداء) ---
scores_C1 <- ind_coords[, "Dim.1"]
ranking <- data.frame(
  الدولة = data$الدولة,
  Score_C1 = round(scores_C1, 3),
  الترتيب = rank(-scores_C1)  # ترتيب تنازلي (الأفضل أولاً)
)
ranking <- ranking[order(ranking$Score_C1, decreasing = TRUE), ]
rownames(ranking) <- NULL

print("ترتيب الدول حسب المؤشر الاقتصادي (المكوّن 1):")
print(ranking)

# إضافة تفسير بسيط
cat("\n📌 التفسير:\n")
cat("- القيم الموجبة على C1 تضخم منخفض، بطالة منخفضة، تشير إلى: GDP عالٍ، استثمار عالٍ.\n")
cat("- القيم السالبة تعكس الوضع المعاكس.\n")
```

باستخدام بايثون:

```python
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn.decomposition import PCA
from sklearn.preprocessing import StandardScaler

# --- 1. إدخال البيانات ---
data = pd.DataFrame({
    'الدولة': ['A', 'B', 'C', 'D', 'E'],
    'معدل_البطالة_%': [12, 8, 25, 20, 6],
    'التضخم_%': [6, 2, 18, 14, 1],
    'GDP_للفرد': [45000, 38000, 8000, 12000, 52000],
    'الاستثمار_في_التعليم_%GDP': [6.2, 5.5, 3.1, 4.0, 7.0]
})

print("البيانات الأصلية:")
print(data)
print("\n" + "="*50 + "\n")

# --- 2. إعداد المتغيرات الكمية ---
X = data.drop('الدولة', axis=1)
countries = data['الدولة'].values

# --- 3. توحيد البيانات ---
scaler = StandardScaler()
X_scaled = scaler.fit_transform(X)
X_scaled_df = pd.DataFrame(X_scaled, columns=X.columns, index=countries)
```

```
print("البيانات بعد التوحيد (Z-scores):")
print(X_scaled_df.round(3))
print("\n" + "="*50 + "\n")

# --- 4. تطبيق PCA ---
pca = PCA()
pca.fit(X_scaled)

# --- 5. القيم الذاتية والتباين ---
eigenvalues = pca.explained_variance_
explained_var_ratio = pca.explained_variance_ratio_
cumulative_var = np.cumsum(explained_var_ratio)

results_df = pd.DataFrame({
    'المكوّن': [f'C{i+1}' for i in range(len(eigenvalues))],
    'القيمة_الذاتية_λ': np.round(eigenvalues, 3),
    'التباين_المفسّر_%': np.round(explained_var_ratio * 100, 2),
    'التباين_تراكمي_%': np.round(cumulative_var * 100, 2)
})

print("تحليل القيم الذاتية:")
print(results_df)
print("\n" + "="*50 + "\n")

# --- 6. أحمال المتغيرات (Loadings) ---
# Loadings = eigenvectors × sqrt(eigenvalues)
loadings = pca.components_.T * np.sqrt(eigenvalues)
loadings_df = pd.DataFrame(
    loadings,
    columns=[f'C{i+1}' for i in range(loadings.shape[1])],
    index=X.columns
)

print("أحمال المتغيرات (Loadings):")
print(loadings_df.round(3))
print("\n" + "="*50 + "\n")

# --- 7. إحداثيات الدول ---
scores = pca.transform(X_scaled)
scores_df = pd.DataFrame(scores, columns=[f'C{i+1}' for i in
range(scores.shape[1])], index=countries)

print("إحداثيات الدول:")
print(scores_df.round(3))
print("\n" + "="*50 + "\n")

# --- 8. الترتيب حسب C1 ---
ranking = scores_df[['C1']].copy().sort_values('C1', ascending=False)
ranking['الترتيب'] = range(1, len(ranking)+1)
ranking.columns = ['Score_C1', 'الترتيب']
```

```
print("ترتيب الدول حسب C1 (من الأفضل إلى الأسوأ):")
print(ranking.round(3))
print("\n" + "="*60)

# --- 9. رسوم بيانية (بدون seaborn) ---
plt.style.use('default')  # استخدام أسلوب بسيط

fig, axes = plt.subplots(2, 2, figsize=(15, 11))
fig.suptitle('تحليل المكوّنات الرئيسية (PCA)', fontsize=16, fontweight='bold', y=0.98)

# أ) Scree Plot
ax = axes[0, 0]
n_comp = len(explained_var_ratio)
ax.bar(range(1, n_comp+1), explained_var_ratio * 100,
       color='steelblue', alpha=0.7, label='تباين كل مكوّن')
ax.plot(range(1, n_comp+1), cumulative_var * 100,
        'ro-', linewidth=2, label='تباين تراكمي')
ax.axhline(80, color='gray', ls='--', lw=0.8)
ax.set_xlabel('رقم المكوّن')
ax.set_ylabel('نسبة التباين (%)')
ax.set_title('منحنى التباين (Scree Plot)')
ax.legend()
ax.grid(True, linestyle=':', alpha=0.6)

# ب) Correlation Circle (C1 × C2)
ax = axes[0, 1]
# دائرة الوحدة
circle = plt.Circle((0, 0), 1, color='black', fill=False, ls='--', lw=1)
ax.add_patch(circle)

# رسم المتغيرات كأسهم
for i, var in enumerate(X.columns):
    x, y = loadings[i, 0], loadings[i, 1]
    ax.arrow(0, 0, x, y,
             head_width=0.04, head_length=0.04,
             fc='darkred', ec='darkred', lw=1.5)
    ax.text(x*1.15, y*1.15, var, fontsize=9, fontweight='bold',
            ha='center', va='center', color='darkred')

ax.set_xlim(-1.1, 1.1)
ax.set_ylim(-1.1, 1.1)
ax.set_xlabel(f'C1 ({explained_var_ratio[0]*100:.0f}%)')
ax.set_ylabel(f'C2 ({explained_var_ratio[1]*100:.0f}%)')
ax.set_title('دائرة الارتباط')
ax.axhline(0, color='gray', lw=0.5)
ax.axvline(0, color='gray', lw=0.5)
ax.set_aspect('equal')

# ج) خريطة الدول (C1 × C2)
ax = axes[1, 0]
sc = ax.scatter(scores[:,0], scores[:,1],
```

```python
                c=scores[:,0], cmap='coolwarm', s=100, edgecolor='k')
for i, (x, y) in enumerate(scores):
    ax.text(x+0.03, y+0.03, countries[i], fontweight='bold', fontsize=11)

ax.set_xlabel(f'C1 ({explained_var_ratio[0]*100:.0f}%)')
ax.set_ylabel(f'C2 ({explained_var_ratio[1]*100:.0f}%)')
ax.set_title('خريطة الدول')
ax.axhline(0, color='gray', lw=0.5)
ax.axvline(0, color='gray', lw=0.5)
plt.colorbar(sc, ax=ax, label='قيمة C1')

# د) Biplot مبسط
ax = axes[1, 1]
# الدول
ax.scatter(scores[:,0], scores[:,1], c='lightblue', s=80, edgecolor='k')
for i, (x, y) in enumerate(scores):
    ax.text(x+0.03, y+0.03, countries[i], fontweight='bold')

# المتغيرات (مُضخّمة قليلاً للوضوح)
scale = 1.4
for i, var in enumerate(X.columns):
    x, y = loadings[i, 0]*scale, loadings[i, 1]*scale
    ax.arrow(0, 0, x, y,
             head_width=0.05, head_length=0.05,
             fc='red', ec='red', alpha=0.8, lw=1.2)
    ax.text(x*1.1, y*1.1, var, fontsize=8, color='darkred', fontweight='bold')

ax.set_xlabel('C1')
ax.set_ylabel('C2')
ax.set_title('Biplot: دول + متغيرات')
ax.axhline(0, color='gray', lw=0.5)
ax.axvline(0, color='gray', lw=0.5)

plt.tight_layout()
plt.show()

# --- 10. تفسير تلقائي ---
print("📌 تفسير موجز:")
c1_var = explained_var_ratio[0] * 100
print(f"- C1 يفسّر {c1_var:.1f}% من التباين → بُعد رئيسي واحد كافٍ تقريبًا.")

pos_vars = [col for col in X.columns if loadings_df.loc[col, 'C1'] > 0]
neg_vars = [col for col in X.columns if loadings_df.loc[col, 'C1'] < 0]
print(f"- C1 يرتفع مع: {', '.join(pos_vars)}")
print(f"- C1 ينخفض مع: {', '.join(neg_vars)}")
print("⇒ يمكن تسميته: **مؤشر التنمية الاقتصادية**")

print("\n- ترتيب الدول:")
for i, (country, row) in enumerate(ranking.iterrows(), 1):
                          print(f"  {int(row['الترتيب'])}. {country}: {row['Score_C1']:.2f}")
```

تمارين:

لدينا مجموعة من معطيات المتمثلة في نقاط امتحان لنسع طلبة في خمس مواد: الرياضيات (Mh)، العلوم (Sc) ، انجليزية (En) ، الأدب (Lt) و موسيقى(Mu)، و هي مبينة في الجدول التالي: (كل فرد لديه نفس الوزن أي $\frac{1}{5}$)

الطلبة/المواد	الرياضيات (Mh)	العلوم (Sc)	انجليزية (En)	الأدب (Lt)	موسيقى(Mu)
1	6	6	5	5.5	8
2	8	8	8	8	9
3	6	7	11	9.5	11
4	14.5	14.5	15.5	15	8
5	14	14	12	12	10
6	11	10	5.5	7	13
7	5.5	7	14	11.5	10
8	13	12.5	8.5	9.5	12
9	9	9.5	12.5	12	18

(بالنسبة للحسابات، تقريب النتائج تكون رقمين بعد الفاصلة)

1. ما هو نوع التحليل العاملي الذي يمكن تطبيقه بالنسبة لهذه المعطيات و لماذا؟

بالاعتماد على برنامج SPSS كانت نتائج التحليل للمتغيرات الأولية كالتالي:

Matrice de corrélation[a]

		الرياضيات (Mh)	العلوم (Sc)	انجليزية (En)	الأدب (Lt)	موسيقى(Mu)
Corrélation	الرياضيات (Mh)	1,000	,983	,227	,490	,011
	العلوم (Sc)	,983	1,000	,397	,634	,006
	انجليزية (En)	,227	,397	1,000	,956	,038
	الأدب (Lt)	,490	,634	,956	1,000	,089
	موسيقى(Mu)	,011	,006	,038	,089	1,000
Signification (unilatéral)	الرياضيات (Mh)		,000	,279	,090	,489
	العلوم (Sc)	,000		,145	,033	,494
	انجليزية (En)	,279	,145		,000	,461
	الأدب (Lt)	,090	,033	,000		,410
	موسيقى(Mu)	,489	,494	,461	,410	

a. Déterminant = 5,398E-6

Indice KMO et test de Bartlett

Indice de Kaiser-Meyer-Olkin pour la mesure de la qualité d'échantillonnage.		,330
Test de sphéricité de Bartlett	Khi-carré approx.	66,712
	ddl	10
	Signification	,000

Matrice de covariance des coefficients des composantes

Composante	1	2	3
1	1,000	,000	,000
2	,000	1,000	,000
3	,000	,000	1,000

Méthode d'extraction : Analyse en composantes principales.
Méthode de rotation : Varimax avec normalisation Kaiser.
Scores des composantes.

Variance totale expliquée

Composante	Valeurs propres initiales Total	% de la variance	% cumulé	Sommes extraites du carré des chargements Total	% de la variance	% cumulé	Sommes de rotation du carré des chargements Total	% de la variance	% cumulé
1							2,039	40,779	40,779
2	1,151			1,151			1,954	39,087	79,866
3	,983			,983			1,002	20,046	99,913
4	,004								
5	,000								

Méthode d'extraction : Analyse en composantes principales.

باستخدام XLSTAT تم الحصول على النتائج التالية:

1- النتائج المتعلقة بالمتغيرات :

أ- تشبع المتغيرات:

	D1	D2	D3
الرياضيات (Mh)	0,991	0,129	0,007
العلوم (Sc)	0,952	0,303	-0,004
انجليزية (En)	0,099	0,995	0,005
الأدب (Lt)	0,373	0,925	0,058
موسيقى(Mu)	0,000	0,033	0,999

ب -مساهمة المتغيرات في تكوين

	D1	D2	D3
الرياضيات (Mh)	48,198	0,855	0,004
العلوم (Sc)	44,490	4,691	0,002
انجليزية (En)	0,485	50,610	0,002
الأدب (Lt)	6,827	43,787	0,334
موسيقى(Mu)	0,000	0,056	99,658

ج- جودة التمثيل :

	D1	D2	D3
الرياضيات (Mh)	**0,983**	0,017	0,000
العلوم (Sc)	**0,907**	0,092	0,000
انجليزية (En)	0,010	**0,989**	0,000
الأدب (Lt)	0,139	**0,856**	0,003
موسيقى(Mu)	0,000	0,001	**0,999**

الأسئلة:

2. هل توجد علاقة معنوية بين الرياضيات والعلوم عند مستوى معنوية 5%
3. ماهي العلاقة بين تحصيل الطلبة في مادة الرياضيات واللغة الإنجليزية
4. استنتج القيمة الذاتية الباقية. مع حساب نسبة تباين (نسبة التمثيل) والتراكم للقيم الذاتية
5. حسب نتائج التحليل، ما هو عدد المركبات التي تم استخلاصها؟
6. علق على نتائج التحليل العاملي لمعطيات الدراسة (عبر ذكر مراحل التحليل لعاملي)
7. قم باقتراح تسمية لهذه المركبات الجديدة؟
8. ما هو نوع أسلوب التدوير الذي يمكن تطبيقه في هذا التحليل ولماذا؟

تمرين 2

الجدول التالي يحتوي على **n = 18** سيارة تحددها **P=6** متغيرات (مميزات):

السرعة القصوى	وزن السيارة	عرض السيارة	طول السيارة	قوة المحرك	حجم المحرك
vitesse maximale	poids	largeur	longueur	puissance	Cylindrée
V.MAX	POIDS	LARG	LONG	PUISS	CYL

Modele	CYL	PUISS	LONG	LARG	POIDS	V.MAX
Alfasud TI	1350	79	393	161	870	165
Audi 100	1588	85	468	177	1110	160
Simca 1300	1294	68	424	168	1050	152
Citroen GS Club	1222	59	412	161	930	151
Fiat 132	1585	98	439	164	1105	165
Lancia Beta	1297	82	429	169	1080	160
Peugeot 504	1796	79	449	169	1160	154
Renault 16 TL	1565	55	424	163	1010	140
Renault 30	2664	128	452	173	1320	180
Toyota Corolla	1166	55	399	157	815	140
Alfetta 1.66	1570	109	428	162	1060	175
Princess 1800	1798	82	445	172	1160	158
Datsun 200L	1998	115	469	169	1370	160
Taunus 2000	1993	98	438	170	1080	167
Rancho	1442	80	431	166	1129	144
Mazda 9295	1769	83	440	165	1095	165
Opel Rekord	1979	100	459	173	1120	173
Lada 1300	1294	68	404	161	955	140

1. **قم باجراء تحليل بواسط المركبات الأساسية على البيانات أعلاه باستخدام احد البرامج الإحصائية : spss ; xlstat,R**

الفصل الرابع : التحليل العاملي التناظري

التحليل العاملي التوافقي / التناظري

L’analyse factorielle des correspondances

تمهيد:

يُعد التحليل العاملي التوافقي (Analyse Factorielle des Correspondances - AFC)، المعروف أيضًا بالتحليل العاملي البسيط، من الأساليب الإحصائية الحديثة لتحليل البيانات، وقد تم تطويره في فرنسا من طرف الباحث Jean-Paul Benzécri بجامعة Pierre and Marie Curie University في باريس، وذلك داخل معهد الإحصاء بباريس ومخبر الإحصاء النظري والتطبيقي، خلال الفترة الممتدة بين 1970 و1990.

ويُعتبر AFC تقنية إحصائية تهدف إلى دراسة العلاقة أو التوافق بين متغيرين نوعيين (فئويين)، من خلال تحليل المعلومات المتضمنة في جداول التوافق أو جداول التقاطع. وتتمثل الفكرة الأساسية لهذه الطريقة في اختزال حجم البيانات وتبسيطها عبر تمثيلها في عدد محدود من المحاور العاملية، مما يسهل تفسير العلاقات بين الفئات المختلفة للمتغيرات.

ويُعد التحليل العاملي التوافقي امتدادًا لأسلوب تحليل المركبات الرئيسية(ACP) ، غير أنه يعتمد على مسافة كاي مربع χ^2 بدلًا من الاعتماد على الارتباط او التباين كما هو الحال في تحليل المركبات الرئيسية.

ويتكون هذا الدرس من محورين: محور نظري وآخر تطبيقي. يشمل المحور النظري المفاهيم العامة المتعلقة بالتحليل العاملي للمراسلات اضافة الى شرح مبادئ AFC وكيفية حسابه بالاعتماد على تحليل المركبات الرئيسية.

أما المحور التطبيقي، فيعتمد على استخدام برنامج spss ولغة البرمجة R لتحليل مجموعة بيانات وتطبيق خطوات التحليل العاملي للمراسلات عمليًا، مع تفسير النتائج المتحصل عليها.

1.التعريف والهدف

يُعدّ التحليل العاملي التوافقي (AFC)أسلوبًا إحصائيًا متعدد المتغيرات يهدف إلى دراسة العلاقة بين متغيرَين نوعيَّين (كيفيَّين) من خلال تمثيل بياني ثنائي البُعد يُوضح مدى الترابط بين فئات المتغيرَين. يرتكز هذا الأسلوب على تحليل جدول التقاطع (جدول التوافق)،

فـAFC هي تقنية تهدف أساسا (مثل ACP) الى وصف اقصى قدر معين من المعلومات الموجودة في جدول يحتوي على معطيات معينة (خاصة الوصف يكون على شكل رسم بياني).

لماذا تقابلي؟ : متغيرات رقمية كمية ← ارتباط corrélation → variable numérique

متغيرات نوعية اسمية ← توافق correspondance → variable nominale

بعض أمثلة استعمال التحليل العاملي التو افقي(AFC)

يُستخدم التحليل العاملي التوافقي (AFC) في العديد من المجالات الاقتصادية والاجتماعية والزراعية والطبية، خاصة عند دراسة العلاقات بين المتغيرات النوعية والفئات المختلفة للبيانات. ومن بين أهم تطبيقاته ما يلي:

1. دراسة توزيع الإطارات حسب قطاع النشاط والفئة العمرية

لنفترض أنك قمت بإجراء مسح إحصائي حول توزيع الإطارات في القطاعات الاقتصادية وفق:

- قطاعات النشاط الاقتصادي المجموعة. (i)
- أصناف الأعمار المجموعة (j)

في هذه الحالة يسمح التحليل العاملي التوافقي بدراسة العلاقة بين القطاع الاقتصادي والبنية العمرية للإطارات، كما يساعد على:

1. تحديد القطاعات ذات الطابع الشباني.
2. تحديد القطاعات التي يغلب عليها العمال الأكبر سنًا.
3. الكشف عن القطاعات التي تمتلك نفس البنية العمرية.
4. تمثيل القطاعات والفئات العمرية بيانياً لإظهار أوجه التقارب أو الاختلاف بينها.

وبذلك يوفر AFC رؤية شاملة حول التوافق بين طبيعة النشاط الاقتصادي والتركيب العمري للموارد البشرية.

2. دراسة تأثير المبيدات على الأعشاب الضارة

يمكن استخدام AFC في المجال الزراعي لدراسة العلاقة بين:

- أنواع المبيدات الزراعية.
- أنواع الأعشاب الضارة.

حيث يسمح التحليل بـ:

1. تحديد المبيدات الأكثر فعالية تجاه أنواع معينة من الأعشاب.
2. تصنيف الأعشاب المتشابهة من حيث استجابتها للمبيدات.
3. الكشف عن أنماط التوافق أو الارتباط بين المبيدات والأعشاب الضارة.
4. تمثيل النتائج بيانياً لتسهيل تفسير العلاقات بين المتغيرات.

3. تحليل سلوك المستهلك

يستخدم AFC كثيرًا في بحوث التسويق لدراسة العلاقة بين:

- الفئات العمرية أو الاجتماعية للمستهلكين.
- أنواع المنتجات أو العلامات التجارية المفضلة.

مما يساعد المؤسسات على:

1. تحديد خصائص الزبائن.
2. فهم تفضيلات المستهلكين.
3. توجيه الاستراتيجيات التسويقية بصورة أكثر فعالية.

4. تحليل نتائج الاستبيانات

في الدراسات الاجتماعية واستطلاعات الرأي، يُستخدم AFC لتحليل العلاقة بين:

- إجابات الأفراد.
- الخصائص الديموغرافية مثل الجنس، العمر، المستوى التعليمي أو المهني.

ويسمح ذلك بالكشف عن الفئات المتشابهة في الآراء أو السلوكيات.

5. المجال الطبي والصحي

يمكن استعمال AFC لدراسة العلاقة بين:

- أنواع الأمراض.
- الفئات العمرية أو العوامل الصحية المختلفة.

وذلك بهدف تحديد الفئات الأكثر تعرضًا لبعض الأمراض أو دراسة أنماط الانتشار الصحي داخل المجتمع.

بصفة عامة، يهدف التحليل العاملي التوافقي (AFC) إلى تحديد درجة الاعتمادية dépendance أو التوافق (Correspondance) بين مجموعتين من المعطيات أو بين فئات المتغيرات i وj ، وذلك من خلال دراسة العلاقات الموجودة داخل جداول التوافق وتحليلها إحصائيًا.

ومثل تحليل المركبات الرئيسية(ACP) ، يُعتبر التحليل العاملي التوافقي أسلوبًا وصفيًا واستكشافيًا بالدرجة الأولى، حيث يساهم في تبسيط البيانات واكتشاف البنية العامة للعلاقات بين المتغيرات، دون أن يقدم تفسيرًا سببيًا مباشرًا للظاهرة المدروسة.

وبالتالي، فإن كلًا من AFC وACP يمثلان مرحلة أولية وأساسية في الدراسة الإحصائية، إذ يساعدان الباحث على فهم طبيعة البيانات، والكشف عن الارتباطات والتقاربات بين المتغيرات أو الفئات المختلفة، غير أنهما لا يقدمان إجابة نهائية عن الإشكالية المطروحة أو الفرضيات البحثية، وإنما يمهدان الطريق لتحليلات إحصائية أعمق وأكثر تفسيرًا.

أهداف التحليل العاملي التو افقي(AFC)

1. دراسة العلاقة والتوافق بين المتغيرات النوعية.
2. تحليل الجداول التوافقية واستخراج المعلومات المهمة منها.
3. اختزال حجم البيانات وتقليل عدد الأبعاد.
4. تمثيل البيانات بيانياً لتسهيل تفسيرها.
5. الكشف عن الفئات المتشابهة أو المتقاربة.
6. إبراز البنية الداخلية والعلاقات الخفية داخل البيانات.
7. تسهيل تفسير الجداول الكبيرة والمعقدة.
8. المقارنة بين الصفوف والأعمدة داخل الجدول التوافقي.
9. المساعدة في اتخاذ القرار وفهم الظواهر المدروسة.

أنواع الجداول التي يمكن أن يعالجها التحليل العاملي التو افقي(AFC)

يختص التحليل العاملي التوافقي (AFC) بدراسة وتحليل **الجداول التو افقية (Contingency Tables)**، وهي جداول تحتوي على تكرارات لمتغيرات نوعية (فئوية)، ولا يمكن معالجتها مباشرة باستخدام تحليل المركبات الرئيسية (ACP) الذي يتعامل أساسًا مع البيانات الكمية.

ومن أهم أنواع الجداول التي يعالجها AFC :

1. الجداول ذات البعد الثنائي (Two-way Contingency Tables)

وهي الجداول البسيطة التي تربط بين متغيرين نوعيين فقط، مثل:

- الصفوف: الفئات i
- الأعمدة: الفئات j
- القيم: التكرارات

مثال:

العلاقة بين:

- الجنس (ذكر/أنثى)
- نوع الاستهلاك (مرتفع/منخفض)

2. الجداول متعددة الفئات(Multi-category Tables)

وهي جداول تحتوي على عدة فئات داخل كل متغير، مثل:

- مستويات التعليم (ابتدائي، ثانوي، جامعي)
- فئات الدخل
- فئات العمر

وتسمح AFC بتحليل العلاقات المعقدة بين هذه الفئات.

3. الجداول المتقاطعة الكبيرة(Large Contingency Tables)

وهي جداول تحتوي على عدد كبير من الصفوف والأعمدة، مثل:

- قطاعات النشاط الاقتصادي × الفئات العمرية
- المنتجات × تفضيلات المستهلكين
- الأمراض × الفئات السكانية

هنا يظهر دور AFC في:

- تبسيط البيانات
- اختزال الأبعاد
- تمثيل العلاقات بيانيًا

4. الجداول النسبية أو المعيارية

وهي جداول يتم تحويل تكراراتها إلى:

- نسب مئوية
- أو تكرارات نسبية

وذلك بهدف تسهيل المقارنة بين الفئات المختلفة.

لماذا لا تعالجهاACP ؟

لأن تحليل المركبات الرئيسية(ACP) :

- مخصص للمتغيرات الكمية (عددية)
- يعتمد على التباين والانحراف المعياري

بينماAFC :

- مخصص للمتغيرات النوعية
- يعتمد على تكرارات الفئات ومسافة كاي مربع: χ^2

يمكن القول إن AFC هو الأداة المناسبة لتحليل **كل الجداول التو افقية التي تحتوي على بيانات نوعية،** خاصة عندما يكون الهدف هو:

1. دراسة العلاقات بين الفئات
2. اكتشاف التوافق أو الترابط بين المتغيرات
3. تمثيل البيانات في فضاء منخفض الأبعاد بشكل بصري واضح

الجدول التو افقي (Table de contingence)

الجدول التوافقي هو جدول إحصائي يتكون من n صفوف و pأعمدة، حيث يمثل كل صف فئة من فئات المتغير الأول، بينما يمثل كل عمود فئة من فئات المتغير الثاني. ويحتوي تقاطع الصف رقم i مع العمود رقم j على القيمة: n_{ij}

والتي تمثل عدد الأفراد أو المشاهدات التي تمتلك في الوقت نفسه الخاصية i والخاصية j .

وبعبارة أخرى، فإن n_{ij} هو التكرار المشترك (Effectif conjoint) للفئتين i وj، أي عدد العناصر التي تنتمي في آن واحد إلى الفئة i من المتغير الأول، وإلى الفئة j من المتغير الثاني.

ويمكن تمثيل الجدول التوافقي بالشكل العام التالي:

	j_1	j_2	...	j_p	المجموع
i_1	n_{11}	n_{12}	...	n_{1p}	$n_{1.}$
i_2	n_{21}	n_{22}	...	n_{2p}	$n_{2.}$
...	...	...	...	...	...
i_n	n_{n1}	n_{n2}	...	n_{np}	$n_{n.}$
المجموع	$n_{.1}$	$n_{.2}$	...	$n_{.p}$	N

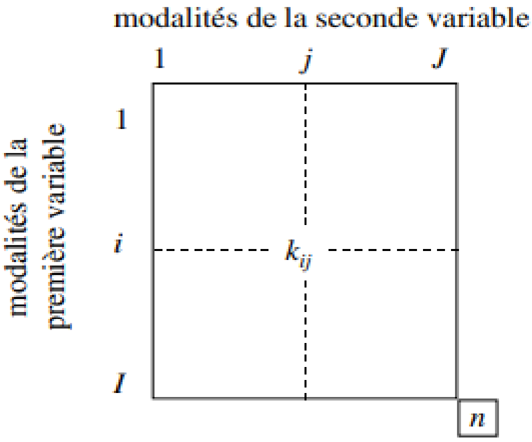
حيث:

$n_{i.}$ يمثل مجموع الصف i .

$n_{.j}$ يمثل مجموع العمود j .

N يمثل الحجم الكلي للعينة : $N = \sum_{i=1}^{n} \sum_{j=1}^{p} n_{ij}$

الجدول 4-1 : جدول البيانات الخامة



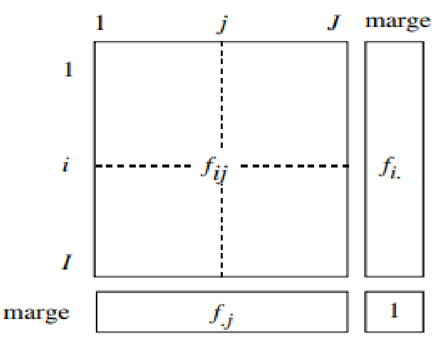
i : مجموعة الصفوف وعددها (8 مستويات تعليمية).

j: مجموعة الأعمدة وعددها (9 فئات مهنية).

k_{ij} : عدد الأفراد الذين يملكون في الوقت نفسه الفئة i من المتغير الأول والفئة j من المتغير الثاني، أي الذين لديهم مستوى التعليم i ويشغلون وظيفة من الفئة j .

مجموع جميع القيم k_{ij} يساوي N، وهو العدد الإجمالي للأفراد.

الجدول 4-2: جدول f للتكرارات النسبية وهوامشه.



$$f_{ij} = k_{ij}/n$$
$$f_{i.} = \sum_j f_{ij}$$
$$f_{.j} = \sum_i f_{ij}$$
$$\sum_i f_{i.} = \sum_j f_{.j} = \sum_i \sum_j f_{ij} = 1$$

خصائص الجدول التو افقي:

1. يحتوي على متغيرين نوعيين أو أكثر.
2. يعتمد على التكرارات (frequencies) وليس القيم العددية المباشرة.
3. يمكن تحويله إلى نسب مئوية أو نسب شرطية.
4. يستخدم أساسًا في التحليل الإحصائي الوصفي والاستكشافي.
5. الافراد لايظهرون بشكل واضح، فهم ممثلون في الجدول على شكل اعداد
6. مكونات الاعمدة والاسطر لها نفس الدور في تحليل الإشكالية قيد الدراسة، يمكن قلب الاعمدة الى أسطر.

ACP : متغيرات /أفراد ، AFC : أسطر/أعمدة

يعتبر الجدول التو افقي البيئة الأساسية التي يعمل عليها التحليل العاملي التو افقي(AFC) ، حيث يقوم هذا الأخير ب:

1. **تحليل التكرارات داخل الجدول**
2. **تحويلها إلى تمثيل بياني**
3. **دراسة التو افق بين الصفوف والأعمدة**

مثال عددي:

المعطيات التالية هي عبارة عن نقاط في مادتين: رياضيات و اللغة الحية في قسم السادسة أساسي:

N°	1	2	3	4	5	6	7	8	9	10	11	12	13	14
رياضيات	9	13	11	10	12	16	18	12	15	18	13	9	17	13
اللغة الحية	9	7	8	10.5	11	12	16.5	9.5	13	16.5	12	3	17	12.5
N°	15	16	17	18	19	20	21	22	23	24	25	26	27	
رياضيات	10	10	15	10	12	10	16	11	10	8	11	14	14	
اللغة الحية	7.5	6.5	13.5	7.5	9	12	17	10	14.5	7.5	12	8.5	11.5	

نقوم بتحديد الفئات في الرياضيات و اللغة الحية :

بما ان الانحراف المعياري في الإنجليزية اكبر منه في الرياضيات، نحدد مثلا :

- 5 فئات في الإنجليزية :]0–4[ ]4–8[ ]8–12[ ]12–16[[20–16]
- 4 فئات في الرياضيات :]0–5[ ]5–10[ ]10–15[[20–15]

ومنه الجدول الكلي التوافقي هو كالتالي:

		CM1	CM2	CM3	CM4	
	فئات الرياضيات \فئات اللغة الحية	[0-5[	[5-10[	[10-15[	[15-20]	المجموع
CA1	[0-4[	0	1	0	0	$n_{1.} = 1$
CA2	[4-8[	0	1	4	0	$n_{2.} = 5$
CA3	[8-12[	0	1	8	0	$n_{3.} = 9$
CA4	[12-16[	0	0	5	3	$n_{4.} = 8$
CA5	[16-20]	0	0	0	4	$n_{5.} = 4$
	المجموع	$n_{.1} = 0$	$n_{.2} = 3$	$n_{.3} = 17$	$n_{.4} = 7$	$N = n_{..} = 27$

ملاحظة : الفئة الأولى في الرياضيات تحتوي على 0 فرد، لهذا يمكن الغاؤها.

يبين لنا هذا الجدول مثلا :

- ان 5 تلاميذ تحصلوا في الرياضيات على نقطة]10-15]وفي الإنجليزية [12-16[في ان واحد.
- ان 8 ثلاميذ تحصلوا في الإنجليزية على تقطة محصورة بين 12 و 16 (دون اخد النقطة 16 بعين الاعتبار)

الجدول الاستكشافي (Tableau exploratoire) للجدول التو افقي

الجدول الاستكشافي في إطار التحليل العاملي التوافقي (AFC) هو جدول مشتق من الجدول التوافقي الأصلي، ويعتمد على التكرارات النسبية بدل التكرارات المطلقة، بهدف تسهيل تحليل العلاقات بين المتغيرات.

من اجل تطبيق AFC لتوضيح الارتباطات للجداول المزدوجة ، يجب إيجاد مجموعة من الجداول التي تمكننا من توضيح ذلك.

الجداول التكرارات النسبية (كلية ، للأسطر، للأعمدة)

الجدول التكرار النسبي كلي:

يتم المرور من جدول التكررات الى جدول التكرارات النسبية باستعمال العلاقة :

$$f_{ij} = \frac{n_{ij}}{N}$$

فئات الرياضيات \ فئات اللغة الحية	[5-10[	[10-15[	[15-20]	المجموع
[0-4[	1/27	0/27	0/27	**1/27**

فئات الرياضيات \ فئات اللغة الحية	[5-10[	[10-15[	[15-20]	المجموع
[4-8[	1/27	4/27	0/27	**5/27**
[8-12[	1/27	8/27	0/27	**9/27**
[12-16[	0/27	5/27	3/27	**8/27**
[16-20]	0/27	0/27	4/27	**4/27**
المجموع	**3/27**	**17/27**	**7/27**	**27/27=1**

الجدول التكرار النسبي للأسطر:(Profils lignes)

في التحليل العاملي التوافقي ((AFC، يمثل الجدول التكراري النسبي للأسطر نسب توزيع كل صف عبر الأعمدة، ويُحسب كما يلي:

$$p_{ij} = \frac{n_{ij}}{n_{i.}}$$

حيث:

- n_{ij}: التكرار داخل الخلية
- $n_{i.}$: مجموع السطر i

فئات اللغة الحية \ فئات الرياضيات	[5-10[	[10-15[	[15-20]	المجموع
[0-4[	1/1 = 1	0/1 = 0	0/1 = 0	**1**
[4-8[	1/5 = 0.20	4/5 = 0.80	0/5 = 0	**1**
[8-12[	1/9 ≈ 0.11	8/9 ≈ 0.89	0/9 = 0	**1**
[12-16[	0/8 = 0	5/8 = 0.625	3/8 = 0.375	**1**
[16-20]	0/4 = 0	0/4 = 0	4/4 = 1	**1**

ملاحظات مهمة

1. كل سطر يمثل توزيع فئة اللغة الحية عبر فئات الرياضيات.
2. مجموع كل سطر يساوي دائمًا: $\sum_j p_{ij} = 1$

الجدول التكراري النسبي للأعمدة(Profils colonnes)

في التحليل العاملي التوافقي (AFC)، يمثل الجدول التكراري النسبي للأعمدة توزيع كل عمود عبر الأسطر، ويُحسب كما يلي:

$$p_{ij} = \frac{n_{ij}}{n_{.j}}$$

حيث:

- n_{ij} : تكرار الخلية
- $n_{.j}$: مجموع العمود j

فئات اللغة الحية \ فئات الرياضيات	[5-10[	[10-15[	[15-20]
[0-4[	1/3 = 0.333	0/17 = 0	0/7 = 0
[4-8[	1/3 = 0.333	4/17 = 0.235	0/7 = 0
[8-12[	1/3 = 0.333	8/17 = 0.471	0/7 = 0
[12-16[	0/3 = 0	5/17 = 0.294	3/7 = 0.429
[16-20]	0/3 = 0	0/17 = 0	4/7 = 0.571

- كل عمود يمثل توزيع فئات اللغة الحية داخل فئة الرياضيات.
- مجموع كل عمود يساوي: $\sum_i p_{ij} = 1$

مؤشر التجاذب/التنافر:

يعد مؤشر التجاذب والتنافر (Indice d'attraction / répulsion) من أهم المؤشرات في التحليل العاملي التوافقي، ويستخدم لقياس مدى قوة العلاقة بين فئة من فئات الصفوف وفئة من فئات الأعمدة مقارنة بما هو متوقع في حالة الاستقلالية، و يعطى عادة بالشكل:

$$I_{ij} = \frac{n_{ij}}{e_{ij}}$$

حيث :

n_{ij} : ا لقيمة المشاهدة

eij : القيمة المتوقعة في حالة الاستقلالية، وتحسب بالعلاقة : $e_{ij} = \frac{n_{i.}\, n_{.j}}{N}$

إذا كان :

- $I_{ij}>1$: يوجد تجاذب (ارتباط أقوى من المتوقع) .
- $I_{ij}=1$: يوجد استقلالية.
- $I_{ij}<1$: يوجد تنافر (ارتباط أضعف من المتوقع) .

بالتعويض عن e_{ij} :

$$I_{ij} = \frac{n_{ij} \cdot N}{n_{i.}\, n_{.j}}$$

مؤشر التجاذب/التنافر هو أداة أساسية في AFC تسمح بـ:
مقارنة العلاقة الحقيقية بين الفئات بما هو متوقع نظريًا في حالة الاستقلال، وبالتالي تحديد قوة و اتجاه الارتباط بين المتغيرات النوعية.

ملاحظة : يمكن حساب مؤشر التجاذب/التنافر مباشرة باستعمال التكرارات النسبية بدل التكرارات المطلقة.

لدينا:

$$f_{.j} = \frac{n_{.j}}{N}, \quad f_{i.} = \frac{n_{i.}}{N}, \quad f_{ij} = \frac{n_{ij}}{N}$$

والمؤشر يصبح:

$$\boxed{I_{ij} = \frac{f_{ij}}{f_{i.}\, f_{.j}}}$$

من المثال السابق :

فئات اللغة الحية \ فئات الرياضيات	[5-10[	[10-15[	[15-20]
[0-4[	9.000	0.000	0.000
[4-8[	1.800	1.271	0.000
[8-12[	1.000	1.412	0.000
[12-16[	0.000	0.991	1.446
[16-20]	0.000	0.000	3.857

مثلا :

خلية: [12-16[× [15-20] :

$$f_{ij} = 3/27 \qquad f_{i.} = 8/27 \qquad f_{.j} = 7/27$$

$$I_{ij} = \frac{3/27}{(8/27)(7/27)}$$

$$I_{ij} = \frac{3}{27} \times \frac{27 \times 27}{56}$$

$$I_{ij} \approx 1.446$$

- أقوى تجاذب في البيانات : [0-4[مع [5-10[: 9.00
- أقوى ارتباط إيجابي : [16-20] مع [15-20] : 3.85
- وجود تنافر في بعض الخلايا بسبب القيم المنخفضة مقارنة بالقيم المتوقعة

اختبار كاي مربع (χ^2) في التحليل العاملي التو افقي AFC :

يعتبر اختبار كاي مربع أساس التحليل العاملي التوافقي(AFC) ، لأنه يسمح بقياس درجة الارتباط أو التوافق بين متغيرين نوعيين انطلاقًا من الجدول التوافقي.

يقارن الاختبار بين:

1. التكرارات المشاهدة N_{ij}
2. والتكرارات المتوقعة E_{ij} في حالة الاستقلالية

إذا كان الفرق كبيرًا فنستنتج بوجود علاقة بين المتغيرين.

وتحسب كاي مربع المحسوبة ($\chi^2_{calculé}$) بالعلاقة التالية :

$$d^2 = \sum_{i=1}^{m_1}\sum_{j=1}^{m_2} \frac{\left(n_{ij} - \frac{n_{i\cdot}n_{\cdot j}}{n}\right)^2}{\frac{n_{i\cdot}n_{\cdot j}}{n}} = n\left[\sum_{i=1}^{m_1}\sum_{j=1}^{m_2} \frac{n_{ij}^2}{n_{i\cdot}n_{\cdot j}} - 1\right].$$

يُستخدم إحصاء كاي مربع لقياس درجة الابتعاد عن حالة الاستقلال بين متغيرين نوعيين، وبالتالي تقييم قوة العلاقة أو التوافق بينهما.

إذا كانت المتغيرات مستقلة تمامًا فإن: $\chi^2 = 0$

أما كلما ازدادت قيمة χ^2 دل ذلك على وجود علاقة أقوى بين المتغيرين

تستخرج قيم كاي مربع من جداول إحصائية معدّة مسبقًا، خاصة عندما يكون عدد درجات الحرية:

$$\ell \leq 30$$

حيث:

- يمثل سطر الجدول عدد درجات الحرية ℓ .
- تمثل الأعمدة الاحتمال التراكمي : $P(\chi^2_\ell > D^2_c)$
- أما القيمة الموجودة داخل الجدول فهي القيمة الحرجة : D^2_c والتي تُقارن مع القيمة المحسوبة لاتخاذ القرار الإحصائي.

عندما يكون:

$$\ell > 30$$

فإن توزيع كاي مربع يقترب من التوزيع الطبيعي، ويمكن استعمال التقريب التالي : $\sqrt{2\chi^2_\ell} - \sqrt{2\ell - 1}$

أي أن إحصائية كاي مربع تصبح تقريبًا موزعة وفق التوزيع الطبيعي المعياري $N(0,1)$

❖ **استعمال القيمة الاحتمالية(p-value)**

مع تطور البرمجيات الإحصائية، أصبح بالإمكان حساب القيمة الاحتمالية مباشرة:

$$P\left(\chi^2_{(m_1-1)(m_2-1)} > D^2\right)$$

مما يجعل استعمال الجداول الإحصائية التقليدية أقل ضرورة.

إذا كانت :

$$P\text{-value} < \alpha$$

فإننا:

- نرفض فرضية الاستقلال
- ونستنتج وجود علاقة معنوية بين المتغيرين.

أما إذا كانت:

$$P\text{-value} > \alpha$$

- فنقبل فرضية الاستقلالية.

تطبيق عددي: من المثال السابق:

فرضيات اختبار كاي مربع:

$$\begin{cases} H_0 : \text{لا توجد علاقة بين نقاط الرياضيات ونقاط اللغة الحية (استقلالية إحصائية)} \\ H_1 : \text{توجد علاقة معنوية بين نقاط الرياضيات ونقاط اللغة الحية (ارتباط إحصائي)} \end{cases}$$

حساب كاي تربيع : لدينا :

البيانات:

- **المجموع الكلي** **n = 27**
- **هوامش الصفوف($n_{i.}$)** : $n_{1.} = 1, n_{2.} = 5, n_{3.} = 9, n_{4.} = 8, n_{5.} = 4$
- **هوامش الأعمدة($n_{.j}$)** : $n_{.1} = 0, n_{.2} = 3, n_{.3} = 17, n_{.4} = 7$

حساب $\frac{n_{ij}^2}{n_{i.}n_{.j}}$ **لكل خلية غير صفرية:**

	[5-10[($n_{.j}$=3)	**[10-15[($n_{.j}$=17)**	**[15-20] ($n_{.j}$=7)**
[0-4[($n_{i.}$=1)	$1^2/(1\times3)$ = **0.3333**	$0^2/(1\times17)$ = **0**	$0^2/(1\times7)$ = **0**
[4-8[($n_{i.}$=5)	$1^2/(5\times3)$ = **0.0667**	$4^2/(5\times17)$ = **0.1882**	$0^2/(5\times7)$ = **0**
[8-12[($n_{i.}$=9)	$1^2/(9\times3)$ = **0.0370**	$8^2/(9\times17)$ = **0.4183**	$0^2/(9\times7)$ = **0**
[12-16[($n_{i.}$=8)	$0^2/(8\times3)$ = **0**	$5^2/(8\times17)$ = **0.1838**	$3^2/(8\times7)$ = **0.1607**
[16-20] ($n_{i.}$=4)	$0^2/(4\times3)$ = **0**	$0^2/(4\times17)$ = **0**	$4^2/(4\times7)$ = **0.5714**

مجموع الحدود:

$$\sum \frac{n_{ij}^2}{n_{i.} \cdot n_{.j}} = 0.3333 + 0.0667 + 0.1882 + 0.0370 + 0.4183 + 0.1838 + 0.1607 + 0.5714 = 1.9594$$

تطبيق القانون: $d^2 = 27 \times [1.9594 - 1] = 27 \times 0.9594$

$$\boxed{d^2 \approx 25.90}$$

درجات الحرية:

$$ddl = (m_1 - 1)(m_2 - 1) = (5-1)(3-1) = 4 \times 2 = 8$$

مقارنة بالقيم الجدولية:

مستوى الدلالة α	القيمة الجدولية $\chi^2_{(8)}$	القرار
5%	15.507	**رفض H_0 ← 25.90 > 15.507**
1%	20.090	**رفض H_0 ← 25.90 > 20.090**

التعليق:

بما أن: $d^2=25.90 > \chi^2_{(8,1\%)}=20.090$،

فإننا **نرفض فرضية الاستقلالية H0 عند مستوى دلالة 1%**، أي بدرجة ثقة تبلغ 99%.

هذا يعني أن العلاقة بين نقاط الرياضيات ونقاط اللغة الحية **ذات دلالة إحصائية قوية جدًا**، ولا يمكن إرجاعها إلى الصدفة.

وقد أكد ذلك جدول مؤشر التجاذب/التنافر الذي كشف عن ارتباط موجب واضح: الطلاب ذوو النقاط المرتفعة في اللغة الحية يميلون إلى النقاط المرتفعة في الرياضيات (تجاذب قوي عند A = 3.86 للفئة [16-20 /] [15-20]) ، والعكس صحيح للفئات الضعيفة (تجاذب عند A = 9 للفئة [0-4] / [5-10]).

خلاصة : تبرر هذه النتيجة الانتقال إلى المرحلة التالية من التحليل العاملي التوافقي (AFC) لتحديد طبيعة هذا الارتباط ومحاوره الرئيسية بيانيًا.

ملاحظة: يعد التحليل العاملي التو افقي (AFC) أسلوبًا استكشافيًا وصفيًا بالدرجة الأولى، وعليه فإن دلالة اختباركاي مربع ليست شرطًا ضروريًا لتطبيقه؛ إذ يمكن أن يكشف AFC عن بنية ارتباطية ذات معنى تفسيري حتى في غياب المعنوية الإحصائية لاختبار الاستقلالية.

اختيار المقياس la métrique

بالنسبة للأسطر : profile ligne

نعرف المقياس لفضاء شعاعي للسحابة النقطية π_i انطلاقا من مقلوب المصفوفة القطرية D_r كمايلي:

$$D_r := \begin{pmatrix} f_{1\cdot} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & f_{p\cdot} \end{pmatrix} \in \mathcal{M}(p \times p).$$

يتم الحصول على مصفوفة الصفوف X_r بقسمة كل صف i من المصفوفة N على وزنه f_i

$$X_r := D_r^{-1} N = \begin{pmatrix} \frac{f_{11}}{f_{1\cdot}} & \cdots & \frac{f_{1q}}{f_{1\cdot}} \\ \vdots & \ddots & \vdots \\ \frac{f_{p1}}{f_{p\cdot}} & \cdots & \frac{f_{pq}}{f_{p\cdot}} \end{pmatrix} \in \mathcal{M}(p \times q).$$

بالنسبة للسطر رقم i يكون على شكل توزيع تكراري شرطي للمتغير V_2 بشرط أن $V_1 = V_i$، ويكون كما يلي:

$$f_i^{V2} := \left(\frac{f_{i1}}{f_{i\cdot}}, \dots, \frac{f_{iq}}{f_{i\cdot}} \right)^t, \quad i = 1, \dots, p,$$

ويمكن كتابة جدول المقاطع الصفية Le tableau des profils-lignes كمايلي :

$$\mathbf{X} = \mathbf{D}_1^{-1}\mathbf{N} \quad \text{où} \quad x_{ij} = \frac{n_{ij}}{n_{i.}}$$

المقياس: $\mathbf{M} = n\,\mathbf{D}_2^{-1}$ مصفوفة الثقل : $\mathbf{D} = \frac{1}{n}\mathbf{D}_1$

ثقل المقطع الصفي : $\frac{n_{i.}}{n}$

المسافة بين مقطعين صفيين e_i و $e_{i'}$ تحسب بالعلاقة :

$$\|e_i - e_{i'}\|_{\mathrm{M}}^2 = \sum_{j=1}^{m_2} \frac{n}{n_{.j}} \left(\frac{n_{ij}}{n_{i.}} - \frac{n_{i'j}}{n_{i'.}} \right)^2$$

بالنسبة للاعمدة: profile -colonnes

نُعرف المصفوفة القطرية بالنسبة للأعمدة كما يلي:

$$D_c := \begin{pmatrix} f_{\cdot 1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & f_{\cdot q} \end{pmatrix} \in \mathcal{M}(q \times q).$$

يتم الحصول على مصفوفة الأعمدة X_c بقسمة كل عمود j من المصفوفة Nعلى وزنه $f_{\cdot j}$

$$X_c := D_c^{-1}N^t = \begin{pmatrix} \frac{f_{11}}{f_{\cdot 1}} & \cdots & \frac{f_{p1}}{f_{\cdot 1}} \\ \vdots & \ddots & \vdots \\ \frac{f_{1q}}{f_{\cdot q}} & \cdots & \frac{f_{pq}}{f_{\cdot q}} \end{pmatrix} \in \mathcal{M}(q \times p).$$

وبالمثل، لكل فئة من المتغير V_2 المرتبطة بالعمود (profil) ، بالنسبة للعمود رقم j كما يلي:

$$f_j^{V1} := \left(\frac{f_{1j}}{f_{\cdot j}}, ..., \frac{f_{pj}}{f_{\cdot j}}\right)^t, \ j = 1, ..., q,$$

ويمكن كتابة جدول المقاطع للاعمدة Le tableau des profils-colonnes كمايلي :

$$\mathbf{X} = \mathbf{D}_2^{-1}\mathbf{N}' \quad \text{où} \quad x_{ji} = \frac{n_{ij}}{n_{.j}}$$

حيث :

المقياس: $\mathbf{M} = n\,\mathbf{D}_1^{-1}$ مصفوفة الثقل : $\mathbf{D} = \frac{1}{n}\mathbf{D}_2$

ثقل المقطع الاعمدة : $\frac{n_{.j}}{n}$

الخاصية الأساسية — التثاني (الازدواجية)

تتمتع سحابتا النقاط — سحابة المقاطع الصفية وسحابة المقاطع العمودية — بقصور ذاتي كلي متساو، وهو ما يسوغ تمثيلهما معا في مستوى عاملي واحد مشترك:

$$I_{g_\ell} = I_{g_c} = I_g = \varphi^2 = \frac{d^2}{n}$$

من المثال السابق :

لدينا : N = 27, m_1 = 5 lignes, m_2 = 3 colonnes ، نهمل الفئة [0-5[لان ($n_{.1}$ = 0)

$$\mathbf{N} = \begin{pmatrix} 0 & 1 & 0 & \\ 0 & 1 & 4 & 0 \\ 0 & 1 & 8 & 0 \\ 0 & 0 & 5 & 3 \\ 0 & 0 & 0 & 4 \end{pmatrix} \longrightarrow \mathbf{N} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 4 & 0 \\ 1 & 8 & 0 \\ 0 & 5 & 3 \\ 0 & 0 & 4 \end{pmatrix}$$

المصفوفة القطرية للهوامش :

D_1=diag(1, 5, 9, 8, 4) D_2=diag(3, 17, 7)

$$\mathbf{D}_1 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 5 & 0 & 0 & 0 \\ 0 & 0 & 9 & 0 & 0 \\ 0 & 0 & 0 & 8 & 0 \\ 0 & 0 & 0 & 0 & 4 \end{pmatrix} \qquad \mathbf{D}_2 = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 17 & 0 \\ 0 & 0 & 7 \end{pmatrix}$$

المصفوفة العكسية:

$$\mathbf{D}_1^{-1} = \mathrm{diag}\left(1, \frac{1}{5}, \frac{1}{9}, \frac{1}{8}, \frac{1}{4}\right) = \mathrm{diag}(1,\ 0.200,\ 0.111,\ 0.125,\ 0.250)$$

$$\mathbf{D}_2^{-1} = \mathrm{diag}\left(\frac{1}{3}, \frac{1}{17}, \frac{1}{7}\right) = \mathrm{diag}(0.333,\ 0.059,\ 0.143)$$

$$\mathbf{D}_1^{-1} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0.200 & 0 & 0 & 0 \\ 0 & 0 & 0.111 & 0 & 0 \\ 0 & 0 & 0 & 0.125 & 0 \\ 0 & 0 & 0 & 0 & 0.250 \end{pmatrix} \qquad \mathbf{D}_2^{-1} = \begin{pmatrix} 0.333 & 0 & 0 \\ 0 & 0.059 & 0 \\ 0 & 0 & 0.143 \end{pmatrix}$$

المقياس χ^2:

$$\mathbf{M}_\ell = n\,\mathbf{D}_2^{-1} = 27 \times \mathrm{diag}(0.333,\ 0.059,\ 0.143) = \mathrm{diag}(9.000,\ 1.588,\ 3.857)$$

$$\mathbf{M}_c = n\,\mathbf{D}_1^{-1} = 27 \times \mathrm{diag}(1,\ 0.200,\ 0.111,\ 0.125,\ 0.250)$$

$$= \mathrm{diag}(27.000,\ 5.400,\ 3.000,\ 3.375,\ 6.750)$$

$$\mathbf{M}_\ell = n\,\mathbf{D}_2^{-1} = \begin{pmatrix} 9.000 & 0 & 0 \\ 0 & 1.588 & 0 \\ 0 & 0 & 3.857 \end{pmatrix}$$

$$\mathbf{M}_c = n\,\mathbf{D}_1^{-1} = \begin{pmatrix} 27.000 & 0 & 0 & 0 & 0 \\ 0 & 5.400 & 0 & 0 & 0 \\ 0 & 0 & 3.000 & 0 & 0 \\ 0 & 0 & 0 & 3.375 & 0 \\ 0 & 0 & 0 & 0 & 6.750 \end{pmatrix}$$

مصفوفة الثقل :

$$\mathbf{D}_\ell = \frac{1}{n}\mathbf{D}_1 = \frac{1}{27}\mathrm{diag}(1,\ 5,\ 9,\ 8,\ 4) = \mathrm{diag}(0.037,\ 0.185,\ 0.333,\ 0.296,\ 0.148)$$

$$\mathbf{D}_c = \frac{1}{n}\mathbf{D}_2 = \frac{1}{27}\mathrm{diag}(3,\ 17,\ 7) = \mathrm{diag}(0.111,\ 0.630,\ 0.259)$$

$$\mathbf{D}_\ell = \frac{\mathbf{D}_1}{n} = \begin{pmatrix} 0.037 & 0 & 0 & 0 & 0 \\ 0 & 0.185 & 0 & 0 & 0 \\ 0 & 0 & 0.333 & 0 & 0 \\ 0 & 0 & 0 & 0.296 & 0 \\ 0 & 0 & 0 & 0 & 0.148 \end{pmatrix}$$

$$\mathbf{D}_c = \frac{\mathbf{D}_2}{n} = \begin{pmatrix} 0.111 & 0 & 0 \\ 0 & 0.630 & 0 \\ 0 & 0 & 0.259 \end{pmatrix}$$

جدول المقاطع الصفية : $\mathbf{X}_\ell = \mathbf{D}_1^{-1}\mathbf{N}$

	[5-10[	**[10-15[**	**[15-20]**
[0-4[	1/1 = **1.000**	0/1 = **0.000**	0/1 = **0.000**
[4-8[	1/5 = **0.200**	4/5 = **0.800**	0/5 = **0.000**
[8-12[	1/9 = **0.111**	8/9 = **0.889**	0/9 = **0.000**
[12-16[	0/8 = **0.000**	5/8 = **0.625**	3/8 = **0.375**
[16-20]	0/4 = **0.000**	0/4 = **0.000**	4/4 = **1.000**
Centre g_ℓ	**0.111**	**0.630**	**0.259**

$$\mathbf{X}_\ell = \begin{pmatrix} 1.000 & 0.000 & 0.000 \\ 0.200 & 0.800 & 0.000 \\ 0.111 & 0.889 & 0.000 \\ 0.000 & 0.625 & 0.375 \\ 0.000 & 0.000 & 1.000 \end{pmatrix} \qquad \mathbf{g}_\ell = \begin{pmatrix} 0.111 \\ 0.630 \\ 0.259 \end{pmatrix}^{\top}$$

جدول المقاطع العمودية : $\mathbf{X}_c = \mathbf{D}_2^{-1}\mathbf{N}'$

	[0-4[	**[4-8[**	**[8-12[**	**[12-16[**	**[16-20]**
[5-10[	1/3 = **0.333**	1/3 = **0.333**	1/3 = **0.333**	0/3 = **0.000**	0/3 = **0.000**
[10-15[	0/17 = **0.000**	4/17 = **0.235**	8/17 = **0.471**	5/17 = **0.294**	0/17 = **0.000**
[15-20]	0/7 = **0.000**	0/7 = **0.000**	0/7 = **0.000**	3/7 = **0.429**	4/7 = **0.571**
Centre g_c	**0.037**	**0.185**	**0.333**	**0.296**	**0.148**

مصفوفة التباين المشترك :

$$\mathbf{V} = \mathbf{X}_\ell'\,\mathbf{D}_\ell\,\mathbf{X}_\ell - \mathbf{g}_\ell\,\mathbf{g}_\ell'$$

$$\mathbf{X}_c = \begin{pmatrix} 0.333 & 0.333 & 0.333 & 0.000 & 0.000 \\ 0.000 & 0.235 & 0.471 & 0.294 & 0.000 \\ 0.000 & 0.000 & 0.000 & 0.429 & 0.571 \end{pmatrix} \qquad \mathbf{g}_c = \begin{pmatrix} 0.037 \\ 0.185 \\ 0.333 \\ 0.296 \\ 0.148 \end{pmatrix}^{\top}$$

$$= \begin{pmatrix} 0.049 & -0.025 & -0.024 \\ -0.025 & 0.077 & -0.052 \\ -0.024 & -0.052 & 0.076 \end{pmatrix}$$

القصور الكلي : inertie totale

نعرِف فيما يلي صيغة القصور الذاتي الكلي لسحابتَي نقاط المقاطع الصفّية والمقاطع العمودية بالنسبة إلى مركزَي ثقلهما على التوالي بـ:

$$Inertie(X_l/g_l) = \sum_{i=1}^{p} f_{i.} d^2_{\chi^2}(i, g_l)$$
$$Inertie(X_c/g_c) = \sum_{j=1}^{q} f_{.j} d^2_{\chi^2}(j, g_c)$$

القصور الذاتي الكلي لسحابة النقاط كمية إحصائية تقيس مدى الانحراف عن الاستقلالية، وتُعطى بالصيغة التالية:

$$\phi^2 = \frac{1}{n} \sum_{i=1}^{p} \sum_{j=1}^{q} \frac{(n_{ij} - \frac{n_{i.}n_{.j}}{n})^2}{\frac{n_{i.}n_{.j}}{n}}$$

$$\phi^2 = \frac{\chi^2}{n}$$

من المثال السابق :

$$\varphi^2 = \mathrm{tr}(\mathbf{V}\,\mathbf{M}_\ell) = \frac{d^2}{n} = \frac{25.90}{27} \approx \mathbf{0.959}$$

$$\varphi^2 \leq \min(m_1 - 1,\ m_2 - 1) = \min(4,\ 2) = 2$$

عدد القيم الذاتية الغيرصفرية(عدد المركبات المستخلصة):

$$q = \min(m_1 - 1,\ m_2 - 1) = \min(4,\ 2) = \mathbf{2}$$

المسافة بمقياس كاي مربع بين مقطعين صفيين:

يهدف التحليل العاملي التوافقي (AFC) إلى دراسة تشتت النقاط حول مركز ثقل السحابة. ولتحقيق ذلك، يجب اختيار مقياس مناسب من أجل تعريف المسافة بين نقطتين.

في حالة الـAFC، تُستعمل مسافة كاي مربع (χ2)، والتي تسمح بإعطاء وزن أكبر للفئات ذات التكرارات الضعيفة، أما المقياس المرتبط بهذه المسافة فيُسمى مقياس كاي مربع.

تكون المسافة بين مقطعين سطري و عمودي حسب الشكل العام التالي :

$$d^2_{\chi^2}(i,i') = (i-i')^t M_r (i-i') := \|i-i'\|^2_{M_r}$$

$$d^2_{\chi^2}(j,j') = \sum_{i=1}^{p} \frac{1}{f_{i\cdot}} \left(\frac{f_{ij}}{f_{\cdot j}} - \frac{f_{ij'}}{f_{\cdot j'}}\right)^2 = \|j-j'\|^2_{M_c}$$

بصيغة مصفوفية:

$$d^2_{\chi^2}(e_i,\ e_{i'}) = (e_i - e_{i'})^\top \mathbf{M}_\ell\,(e_i - e_{i'})$$

حيث: $\mathbf{M}_\ell = n\,\mathbf{D}_2^{-1}$

$$M_c := D_r^{-1} = \begin{pmatrix} 1/f_{1\cdot} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1/f_{p\cdot} \end{pmatrix} \qquad M_r := D_c^{-1} = \begin{pmatrix} 1/f_{\cdot 1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1/f_{\cdot q} \end{pmatrix}$$

خواصها الأساسية:

1. **التوازن بين الفئات :**الوزن $\frac{n}{n_{\cdot j}}$ يصحح أثر التكرارات الكبيرة، فالفئات النادرة لا تُهمل.
2. **خاصية التكافؤ التوزيعي :**إذا كان لعمودين j و j' نفس المقطع العمودي، فإن دمجهما لا يغير المسافة بين أي مقطعَين صفيين.
3. **الاستقلالية :**في حالة الاستقلالية الإحصائية تنكمش جميع المقاطع الصفّية إلى مركز الثقل g_ℓ، فتصبح المسافة بين أي مقطعَين تساوي صفرًا.

من المثال السابق :

المسافة بين e_1 ([0-4[) و e_5 ([16-20]) :

$$e_1 = \begin{pmatrix} 1.000 \\ 0.000 \\ 0.000 \end{pmatrix} \qquad e_5 = \begin{pmatrix} 0.000 \\ 0.000 \\ 1.000 \end{pmatrix}$$

$$d^2_{\chi^2}(e_1,\ e_5) = \frac{27}{3}(1-0)^2 + \frac{27}{17}(0-0)^2 + \frac{27}{7}(0-1)^2$$

$$= 9.000 \times 1 + 1.588 \times 0 + 3.857 \times 1$$

$$\boxed{d^2_{\chi^2}(e_1,\ e_5) = 12.857}$$

هذه أكبر مسافة ممكنة في جدولنا، مما يعكس التعارض التام بين الفئتين الطرفيَتين.

العوامل الرئيسية والمركبات الرئيسية:Facteurs principaux et Composantes Principales

تسمى العوامل الرئيسية للمقاطع الصفّية (وبالمثل للمقاطع العمودية) المتجهات $w_i := M_r u_i$ او بالنسبة للاعمدة : $\widetilde{w}_i := M_c \widetilde{u}_i$

المركبات الرئيسية هي **إسقاط** نقاط سحابة المقاطع الصفّية على المحور العاملي k، في حين أن العوامل الرئيسية هي **الاتجاهات** التي تُعظِّم القصور الذاتي المُفسر.

العوامل الرئيسية هي القيم الذاتية للمصفوفة : $M_l X_l^t D_l X_l$ حيث :

$$\begin{aligned} M_l X_l^t D_l X_l &= (nD_c^{-1})(D_l^{-1}N)^t \frac{D_l}{n}(D_l^{-1}N) \\ &= D_c^{-1}N^t D_l^{-1}N \end{aligned}$$

والمركبات الرئيسية هي القيم الذاتية للمصفوفة : $X_l M_l X_l^t D_l$ حيث :

$$X_l M_l X_l^t D_l = D_l^{-1} N D_c^{-1} N^t$$

من المثال السابق :

المصفوفة المراد تقطيرها هي: $D_c^{-1}N^t D_l^{-1}N$ او حسب مثالنا : $\mathbf{A} = \mathbf{D}_2^{-1}\mathbf{N}^\top\mathbf{D}_1^{-1}\mathbf{N}$

وهي مصفوفة **(3×3)** لأن: $\min(m_1, m_2) = \min(5,3) = 3$

$$\mathbf{N}^\top\mathbf{D}_1^{-1}\mathbf{N} = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 \\ 0 & 4 & 8 & 5 & 0 \\ 0 & 0 & 0 & 3 & 4 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{5} & 0 \\ 0 & 0 & \frac{1}{9} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 & 0 & 0 \\ 0 & 4 & 8 & 5 & 0 \\ 0 & 0 & 0 & 3 & 4 \end{pmatrix}$$

$$\mathbf{N}^\top \mathbf{D}_1^{-1}\mathbf{N} = \begin{pmatrix} 1\times 1 + 1\times \frac{1}{5} + 1\times \frac{1}{9} & 1\times \frac{4}{5} + 1\times \frac{8}{9} & 0+0+0 \\ \frac{4}{5}+\frac{8}{9} & \frac{16}{5}+\frac{64}{9}+\frac{25}{8} & \frac{15}{8} \\ 0 & \frac{15}{8} & \frac{9}{8}+\frac{16}{4} \end{pmatrix}$$

$$= \begin{pmatrix} 1+0.200+0.111 & 0.800+0.889 & 0 \\ 0.800+0.889 & 3.200+7.111+3.125 & 1.875 \\ 0 & 1.875 & 1.125+4.000 \end{pmatrix}$$

$$= \begin{pmatrix} 1.311 & 1.689 & 0 \\ 1.689 & 13.436 & 1.875 \\ 0 & 1.875 & 5.125 \end{pmatrix}$$

ومنه :

$$\mathbf{A} = \begin{pmatrix} \frac{1}{3} & 0 & 0 \\ 0 & \frac{1}{17} & 0 \\ 0 & 0 & \frac{1}{7} \end{pmatrix} \begin{pmatrix} 1.311 & 1.689 & 0 \\ 1.689 & 13.436 & 1.875 \\ 0 & 1.875 & 5.125 \end{pmatrix}$$

$$\boxed{\mathbf{A} = \begin{pmatrix} 0.437 & 0.563 & 0.000 \\ 0.099 & 0.790 & 0.110 \\ 0.000 & 0.268 & 0.732 \end{pmatrix}}$$

التحقق:

$$\text{tr}(\mathbf{A}) = 0.437 + 0.790 + 0.732 = 1.959$$

$$\sum_{k=1}^{q} \lambda_k = \text{tr}(\mathbf{A}) - 1 = 1.959 - 1 = 0.959 = \varphi^2$$

القيم الذاتية الفعلية للتحليل هي λ_k بحيث:

$$\det(\mathbf{A} - \mu_k \mathbf{I}) = 0 \quad \text{ثم} \quad \lambda_k = \mu_k - 1$$

وعدد القيم الذاتية غير صفرية:

$$q = \min(m_1 - 1,\ m_2 - 1) = \min(4,\ 2) = \mathbf{2}$$

ملاحظة : عمليا من الصعب تقطير المصفوفة A لانها ، عكس مصفوفة الارتباط في ACP، ليست متناظرة

القيم الذاتية والمتجهات الذاتية:

$$\det(\mathbf{A}-\lambda\,\mathbf{I})=0$$

نحل المعادلة المميزة:

$$\det\begin{pmatrix}0.437-\lambda & 0.563 & 0\\ 0.099 & 0.790-\lambda & 0.110\\ 0 & 0.268 & 0.732-\lambda\end{pmatrix}=0$$

$$(0.437-\lambda)\left[(0.790-\lambda)(0.732-\lambda)-0.110\times 0.268\right]-0.563\left[0.099(0.732-\lambda)\right]=0$$

$$(0.437-\lambda)\left[(0.790-\lambda)(0.732-\lambda)-0.02948\right]-0.05569(0.732-\lambda)=0$$

$$(0.790-\lambda)(0.732-\lambda)=0.5783-1.522\lambda+\lambda^2$$

$$\Rightarrow (0.437-\lambda)(\lambda^2-1.522\lambda+0.5488)-0.05569(0.732-\lambda)=0$$

بعد التبسيط نحصل على المعادلة التكعيبية:

$$-\lambda^3+1.959\lambda^2-1.118\lambda+0.1595=0$$

او:

$$\lambda^3-1.959\lambda^2+1.118\lambda-0.1595=0$$

القيمة الذاتية المهملة $\lambda_0=1$ دائمًا موجودة، نقسم:

$$(\lambda-1)(\lambda^2-0.959\lambda+0.1595)=0$$

نحل الحد التربيعي:

$$\lambda^2-0.959\lambda+0.1595=0$$

$$\Delta=(0.959)^2-4\times 0.1595=0.9197-0.6380=0.2817$$

$$\sqrt{\Delta}=0.5307$$

$$\boxed{\lambda_1=\frac{0.959+0.5307}{2}=0.7449} \qquad \boxed{\lambda_2=\frac{0.959-0.5307}{2}=0.2141}$$

التحقق:

$$\lambda_1+\lambda_2=0.7449+0.2141=0.959=\varphi^2$$

$$\lambda_1\times\lambda_2=0.7449\times 0.2141=0.1595$$

2. المتجهات الذاتية للمقاطع الصفّية (العوامل الرئيسية) (u_k)

نحل : $$(\mathbf{A}-\lambda_k\,\mathbf{I})\,\mathbf{u}_k=\mathbf{0}$$

للمحور الأول $\lambda_1=0.7449$:

$$\begin{pmatrix} -0.3079 & 0.563 & 0 \\ 0.099 & 0.0451 & 0.110 \\ 0 & 0.268 & -0.0129 \end{pmatrix} \mathbf{u}_1 = \mathbf{0}$$

من السطر الثالث:

$$0.268\, u_2 - 0.0129\, u_3 = 0 \Rightarrow u_2 = \frac{0.0129}{0.268}\, u_3 = 0.0481\, u_3$$

من السطر الأول:

$$-0.3079\, u_1 + 0.563\, u_2 = 0 \Rightarrow u_1 = \frac{0.563 \times 0.0481}{0.3079}\, u_3 = 0.0880\, u_3$$

نضع $u_3=1$:

$$\mathbf{u}_1 = \begin{pmatrix} 0.0880 \\ 0.0481 \\ 1 \end{pmatrix}$$

للمحور الثاني $\lambda_2=0.2141$: نفس الطريقة ، نحصل على مايلي :

$$\mathbf{u}_2 = \begin{pmatrix} 4.8809 \\ -1.9324 \\ 1 \end{pmatrix}$$

3 . المركبات الرئيسية للمقاطع الصقّية $a_k = X_\ell u_k$

للمحور الأول:

$$\mathbf{a}_1 = \mathbf{X}_\ell\, \mathbf{u}_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0.200 & 0.800 & 0 \\ 0.111 & 0.889 & 0 \\ 0 & 0.625 & 0.375 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0.0880 \\ 0.0481 \\ 1 \end{pmatrix} = \begin{pmatrix} 0.0880 \\ 0.0561 \\ 0.0527 \\ 0.1406 \\ 1.0000 \end{pmatrix}$$

للمحور الثاني:

$$\mathbf{a}_2 = \mathbf{X}_\ell\, \mathbf{u}_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0.200 & 0.800 & 0 \\ 0.111 & 0.889 & 0 \\ 0 & 0.625 & 0.375 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 4.8809 \\ -1.9324 \\ 1 \end{pmatrix} = \begin{pmatrix} 4.8809 \\ 0.4237 \\ -1.1737 \\ -0.8327 \\ 1.0000 \end{pmatrix}$$

4 . علاقة العبور - المركبات الرئيسية للمقاطع العمودية b_k :

$$\boxed{\mathbf{b}_k = \frac{1}{\sqrt{\lambda_k}} \mathbf{D}_2^{-1}\mathbf{N}^\top \mathbf{a}_k \cdot \frac{1}{n}}$$

أو بصيغة مبسطة:

$$b_{jk} = \frac{1}{\sqrt{\lambda_k}} \sum_{i=1}^{m_1} \frac{n_{ij}}{n_{.j}} a_{ik}$$

للمحور الأول: $(\sqrt{\lambda_1} = \sqrt{0.7449} = 0.8631)$

$$b_{1,1} = \frac{1}{0.8631} \times \frac{1}{3}(1 \times 0.0880 + 1 \times 0.0561 + 1 \times 0.0527 + 0 + 0) = \frac{0.0656}{0.8631} = 0.0760$$

$$b_{2,1} = \frac{1}{0.8631} \times \frac{1}{17}(0 + 4 \times 0.0561 + 8 \times 0.0527 + 5 \times 0.1406 + 0) = \frac{0.0848}{0.8631} = 0.0982$$

$$b_{3,1} = \frac{1}{0.8631} \times \frac{1}{7}(0 + 0 + 0 + 3 \times 0.1406 + 4 \times 1.000) = \frac{0.6317}{0.8631} = 0.7319$$

$$\boxed{\mathbf{b}_1 = \begin{pmatrix} 0.0760 \\ 0.0982 \\ 0.7319 \end{pmatrix}}$$

للمحور الثاني $(\sqrt{\lambda_2} = \sqrt{0.2141} = 0.4627)$

$$b_{1,2} = \frac{1}{0.4627} \times \frac{1}{3}(4.8809 + 0.4237 + (-1.1737)) = \frac{1.3770}{0.4627} = 2.9762$$

$$b_{2,2} = \frac{1}{0.4627} \times \frac{1}{17}(0 + 4 \times 0.4237 + 8 \times (-1.1737) + 5 \times (-0.8327)) = \frac{-0.6912}{0.4627} = -$$

$$b_{3,2} = \frac{1}{0.4627} \times \frac{1}{7}(0 + 0 + 0 + 3 \times (-0.8327) + 4 \times 1.000) = \frac{0.2426}{0.4627} = 0.5243$$

$$\boxed{\mathbf{b}_2 = \begin{pmatrix} 2.9762 \\ -1.4939 \\ 0.5243 \end{pmatrix}}$$

5 .الجدول الملخص للإحداثيات:

الفئة	a_{i1}	a_{i2}
[0-4[	0.088	4.881
[4-8[	0.056	0.424
[8-12[	0.053	−1.174
[12-16[	0.141	−0.833
[16-20]	1.000	1.000

الفئة	b_{j1}	b_{j2}
[5-10[	0.076	2.976
[10-15[	0.098	−1.494
[15-20]	0.732	0.524

المحور الأول يفسر **77.7%** من القصور الذاتي الكلي، والمحوران معا يفسران **100%**

المساهمة في القصور الذاتي (CTR) contribution dans l'inertie

تقيس مساهمة الفئة في المحور العاملي مدى إسهامها في بناء ذلك المحور، وتحسب بالعلاقة:

$$CTR(i,k) = \frac{n_{i.}}{n} \cdot \frac{a_{ik}^2}{\lambda_k}$$

مساهمة الفئة i في المحور k

وتعد الفئة مُؤسِّسة للمحور إذا كانت مساهمتها أكبر من وزنها الهامشي $\frac{n_{i.}}{n}$

جودة التمثيل على المستوى العاملي $\cos^2$

تقيس جودة التمثيل مدى دقة إسقاط الفئة على المستوى العاملي المختار، وتُعطى بمربع جيب تمام الزاوية بين المتجه والمستوى:

$$cos^2\theta(i,k^*) = \frac{\sum_{k=1}^{k^*} a_{ik}^2}{\sum_{k=1}^{q} a_{ik}^2}$$

جودة التمثيل ($cos^2\theta$) للفئة i على المحاور الأولى k^*

وتكون الفئة مُمثلة جيدا إذا كان $\cos^2\theta > 0.8$، ومُمثَّلة ضعيفا إذا كان $\cos^2\theta < 0.5$، مما يستوجب الحذر عند تفسير موقعها في المخطط العاملي.

من المثال السابق:

➢ **بالنسبة للاسطر:**

الأوزان : $\frac{n_{i.}}{n}$ = (0.037, 0.185, 0.333, 0.296, 0.148)

القيم الذاتية : λ_1=0.7449 ، λ_2=0.2141

أ) مساهمة المقاطع الصفّية CTR :

$$CTR(i,k) = \frac{n_{i.}}{n} \cdot \frac{a_{ik}^2}{\lambda_k}$$

الفئة	$\frac{n_{i.}}{n}$	a_{i1}	a_{i2}	$CTR(i,1)$	$CTR(i,2)$
[0-4[	0.037	0.088	4.881	$0.037 \times \frac{0.0077}{0.7449} = \mathbf{0.0004}$	$0.037 \times \frac{23.824}{0.2141} = \mathbf{4.117}$
[4-8[	0.185	0.056	0.424	$0.185 \times \frac{0.0031}{0.7449} = \mathbf{0.0008}$	$0.185 \times \frac{0.1798}{0.2141} = \mathbf{0.155}$
[8-12[	0.333	0.053	−1.174	$0.333 \times \frac{0.0028}{0.7449} = \mathbf{0.0013}$	$0.333 \times \frac{1.3783}{0.2141} = \mathbf{2.144}$
[12-16[	0.296	0.141	−0.833	$0.296 \times \frac{0.0199}{0.7449} = \mathbf{0.0079}$	$0.296 \times \frac{0.6939}{0.2141} = \mathbf{0.959}$
[16-20]	0.148	1.000	1.000	$0.148 \times \frac{1.000}{0.7449} = \mathbf{0.1987}$	$0.148 \times \frac{1.000}{0.2141} = \mathbf{0.6912}$
المجموع	**1.000**			**≈ 0.208**	**≈ 8.066**

ب) جودة التمثيل $\cos^2$ للمقاطع الصفّية:

$$cos^2\theta(i) = \frac{a_{i1}^2 + a_{i2}^2}{a_{i1}^2 + a_{i2}^2} = 1 \quad (\text{محوران} = 100\% \text{ دائمًا})$$

لكن على كل محور منفردا:

$$cos^2\theta(i,1) = \frac{a_{i1}^2}{a_{i1}^2 + a_{i2}^2} \qquad cos^2\theta(i,2) = \frac{a_{i2}^2}{a_{i1}^2 + a_{i2}^2}$$

الفئة	a_{i1}^2	a_{i2}^2	$\sum a_{ik}^2$	$cos^2(i,1)$	$cos^2(i,2)$
[0-4[	0.0077	23.824	23.832	**0.000**	**1.000**
[4-8[	0.0031	0.1798	0.183	**0.017**	**0.983**
[8-12[	0.0028	1.3783	1.381	**0.002**	**0.998**
[12-16[	0.0199	0.6939	0.714	**0.028**	**0.972**
[16-20]	1.000	1.000	2.000	**0.500**	**0.500**

➢ بالنسبة للاعمدة:

الأوزان : $\frac{n_{.j}}{n} = (0.111, 0.630, 0.259)$

أ) مساهمة المقاطع العمودية CTR :

الفئة	$\frac{n_{.j}}{n}$	b_{j1}	b_{j2}	$CTR(j,1)$	$CTR(j,2)$
[5-10[	0.111	0.076	2.976	$0.111 \times \frac{0.0058}{0.7449} = \mathbf{0.0008}$	$0.111 \times \frac{8.857}{0.2141} = \mathbf{4.594}$
[10-15[	0.630	0.098	−1.494	$0.630 \times \frac{0.0096}{0.7449} = \mathbf{0.0081}$	$0.630 \times \frac{2.232}{0.2141} = \mathbf{6.569}$
[15-20]	0.259	0.732	0.524	$0.259 \times \frac{0.5358}{0.7449} = \mathbf{0.1864}$	$0.259 \times \frac{0.2746}{0.2141} = \mathbf{0.332}$
المجموع	**1.000**			**≈ 0.195**	**≈ 11.495**

ب) جودة التمثيل $\cos^2$ للمقاطع العمودية:

الفئة	b_{j1}^2	b_{j2}^2	$\sum b_{jk}^2$	$cos^2(j,1)$	$cos^2(j,2)$
[5-10[	0.0058	8.857	8.863	**0.001**	**0.999**
[10-15[	0.0096	2.232	2.242	**0.004**	**0.996**
[15-20]	0.5358	0.2746	0.810	**0.661**	**0.339**

يمكن تلخيص ما سبق في الجدول التالي :

	المحور الأول (77.7%)	المحور الثاني (22.3%)
أكثر المقاطع الصفية مساهمةً	[16-20]	[0-4[و[8-12[
أكثر المقاطع العمودية مساهمةً	[15-20]	[5-10[و[10-15[
أحسن المُمثَّلين	[16-20] و[15-20]	جميع الفئات عدا [16-20]

5. التعليق

1. المحور الأول يميز بوضوح الفئة [16-20] في مقابل باقي الفئات، ويرتبط بفئة الرياضيات [15-20]، مما يعكس التميُّز الدراسي المزدوج.
2. المحور الثاني يبرز الفئة [0-4[في مقابل [8-12[و**[12-16[، مرتبطًا بفئتي [5-10[و **[10-15[في الرياضيات.

3. الفئة [16-20]مُمثلة بالتساوي على المحورَين (cos^2=0.5) مما يعني أنها تُسهم في كليهما.
4. باقي الفئات الصفّية مُمثَّلة بشكل شبه كلي على المحور الثاني .(cos^2≈1)

خلاصة التحليل:

1 .**جودة التمثيل الكلي:** يتوفر لدينا محوران عاملياً فقط(q = 2) ، وهما يُفسّران 100% من القصور الذاتي الكلي:

المحور	القيمة الذاتية	نسبة القصور الذاتي	التراكمي
المحور 1	0.7449	77.7%	77.7%
المحور 2	0.2141	22.3%	100%

المستوى العاملي الأول (المحور 1 + المحور 2) يعيد تمثيل البنية الكاملة للارتباط بين المتغيرَين دون أي فقدان في المعلومة.

2 .**تفسير المحور الأول(77.7%) :** يهيمن على هذا المحور:

- من جهة المقاطع الصفّية :الفئة [16-20] بمساهمة 0.1987 وهي الأعلى بفارق كبير
- من جهة المقاطع العمودية :الفئة [15-20] بمساهمة 0.1864

كلتا الفئتَين تقعان في نفس الجانب الموجب من المحور، مما يكشف عن ترابط قوي بين التميز في اللغة الحية والتميز في الرياضيات .يُعبر المحور الأول إذن عن مستوى التحصيل الدراسي العام.

3 .**تفسير المحور الثاني(22.3%)** : يُهيمن على هذا المحور:

- من جهة المقاطع الصفّية :الفئة [0-4]](إحداثية موجبة كبيرة = +4.881) في مقابل]8-12[(إحداثية سالبة = −1.174) و**[12-16]** (إحداثية سالبة = −0.833)
- من جهة المقاطع العمودية :]5-10[(موجبة = +2.976) في مقابل]10-15[(سالبة = −1.494)

يُعبِّر المحور الثاني عن التعارض بين الطلاب الضعيفين جدا في اللغة الحية ([0-4]) والطلاب المتوسطين ([8-12[و[12-16[)، مرتبطًا بتعارض مماثل بين فئتَي الرياضيات [5-10[و[10-15[.

4 .**قراءة مؤشر التجاذب والتنافر:** تؤكد قراءة المخطط العاملي ما كشفه مؤشر التجاذب/التنافر:

الارتباط	التفسير
[16-20] ↔ [15-20]	الطلاب الممتازون في كلا المادتَين — (A=3.86) تجاذب قوي

الارتباط	التفسير
[0-4[↔ [5-10[	الطلاب الضعيفون في كلا المادتَين — (A=9.00) تجاذب قوي
[16-20] ↔ [5-10[	لا تجمعهما علاقة — (A=0.00) تنافر تام

5. جودة تمثيل الفئات

- جميع المقاطع الصفّية والعمودية مُمثَّلة تمثيلًا ممتازًا ($cos^2≈1$) باستثناء **[16-20]** التي تتوزع مساهمتها بالتساوي بين المحورَين ($cos^2=0.5$ لكل محور)، مما يعني أنها فئة ذات خصوصية مزدوجة.

النتيجة العامة : تكشف نتائج التحليل العاملي التو افقي عن ارتباط موجب قوي ومعنوي بين نقاط الرياضيات ونقاط اللغة الحية) مؤكَّد بـ $\chi^2=25.90$ عند مستوى دلالة 1 % . (يسير الطلاب في كلا المادتَين باتجاه واحد: الطلاب المتميزون في اللغة الحية هم أنفسهم المتميزون في الرياضيات، والطلاب الضعيفون في إحداهما يميلون إلى الضعف في الأخرى، مما يُرجِّح وجود قدرة معرفية عامة مشتركة تؤثر في أداء الطالب في المادتَين معًا.

أسئلة نظرية :

1. متى نستخدم تقنية التحليل العاملي التوافقي **AFC**؟
2. اذكر مثالا يمكن ان نطبق عليه طريقة **AFC**
3. هل يمكن تطبيق **AFC** على الجداول الكمية (متغيرات/مشاهدات)؟ مع الشرح
4. ما هو الفرق بين طريقة **ACP** وطريقة **AFC** ؟

تمارين :

التمرين الأول :

قام جراح بإجراء عمليات جراحية على اليد، لعينة من المرضى تتراوح أعمارهم ما بين **20** و **70** سنة، و قام بتقييم نتائج هذه العمليات بعبارات: "**سيئ**" ، "**متوسط**"، "**حسن**"، و "**ممتاز**"، حسب نتائج كل عملية جراحية على كل مريض.

تم تحديد عمر المرضى في فئات حسب مركز الفئة حيث أن طول كل الفئة 10، فمثلا فئة العمرية من **20** إلى **30** سنة تكتب **C25**، ينتج لنا إذا **05** فئات مبينة في الجدول التالي :

عدد المرضى	الفئة العمرية للمرضى	نتائج العمليات
52	C45	حسن
150	C45	ممتاز
05	C55	سيئ
07	C55	متوسط
33	C55	حسن
89	C55	ممتاز
03	C65	سيئ
10	C65	متوسط
39	C65	حسن
63	C65	ممتاز

عدد المرضى	الفئة العمرية للمرضى	نتائج العمليات
01	C25	سيئ
00	C25	متوسط
06	C25	حسن
16	C25	ممتاز
04	C35	سيئ
04	C35	متوسط
27	C35	حسن
59	C35	ممتاز
06	C45	سيئ
11	C45	متوسط

المطلوب:

1. ما هو عدد متغيرات الدراسة مع ذكرها؟
2. قم بإنشاء جدول المزدوج الكلي.
3. قم بإنشاء جدول الاستكشافي الكلي (التقريب يكون رقمين بعد الفاصلة).
4. قم بحساب مؤشر التجاذب و التنافر لـ $\mathbf{d}_{13}$ و $\mathbf{d}_{22}$ مع التعليق
5. هل هناك علاقة بين عمر المرضى والنتيجة المتعلقة بالعمليات الجراحية؟
6. ما هو عدد المحاور التي يتم استخلاصها عند اجراء تحليل **AFC**؟ ولماذا؟

مثال تطبيقي :

انطلاقا من دراسة أجراها المعهد الوطني للإحصاء حول توزيع المؤسسات الاقتصادية، تحصلنا على الجدول التالي الذي يبين توزيع هذه المؤسسات على بعض الولايات وبعض القطاعات الاقتصادية:

عدد المؤسسات	القطاعات	الولايات
1912	الصناعة	عنابة
2452	الصناعة	الجزائر
5212	الصناعة	وهران
9071	الصناعة	تلمسان
2021	الصناعة	سطيف
258	البناء	عنابة
644	البناء	الجزائر
185	البناء	وهران
171	البناء	تلمسان
2230	البناء	سطيف

عدد المؤسسات	القطاعات	الولايات
10519	التجارة	عنابة
16175	التجارة	الجزائر
51618	التجارة	وهران
17549	التجارة	تلمسان
15888	التجارة	سطيف
7210	الخدمات	عنابة
62734	الخدمات	الجزائر
16297	الخدمات	وهران
9472	الخدمات	تلمسان
15996	الخدمات	سطيف

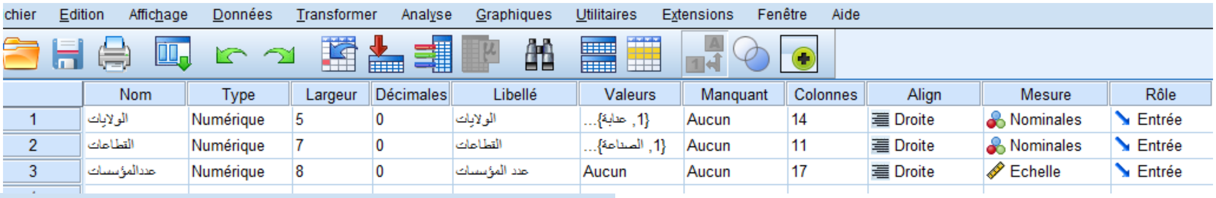
المصدر:عن الديوان الوطني للإحصاء، التعداد الاقتصادي الأول 2011

اعداد الجدول التو افقي/ المزدوج:

	الصناعة	البناء	التجارة	الخدمات	**marge**
عنابة	1912	258	30519	7210	**39899**
الجزائر	2452	644	16175	62734	**82005**
وهران	5212	185	51618	16297	**73312**
تلمسان	9071	171	17549	9472	**36263**
سطيف	2021	2230	15888	15996	**36135**
marge	**20668**	**3488**	**131749**	**111709**	**267614**

ملاحظة : بالنسبة لبرنامج spss، لا يمكن استخدام الجدول المزدوج مباشرة ، و انما نستعمل الجدول الأول : افراد/ مشاهدات ، ثم نحوله الى جدول مزدوج عن طريق ترجيح الاوزان بالنسبة للتكرارات (عدد المؤسسات)

- لدينا متغيران : الولايات و هو متغير نوعي اسمي و القطاعات الاقتصادية متغير نوعي اسمي ، إضافة الى التكرارات المتمثلة في عدد المؤسسات وهو عبارة عن متغير كمي يستعمل لترجيح الاوزان.



	Nom	Type	Largeur	Décimales	Libellé	Valeurs	Manquant	Colonnes	Align	Mesure	Rôle
1	الولايات	Numérique	5	0	الولايات	{1, عنابة}...	Aucun	14	Droite	Nominales	Entrée
2	القطاعات	Numérique	7	0	القطاعات	{1, الصناعة}...	Aucun	11	Droite	Nominales	Entrée
3	عددالمؤسسات	Numérique	8	0	عدد المؤسسات	Aucun	Aucun	17	Droite	Echelle	Entrée

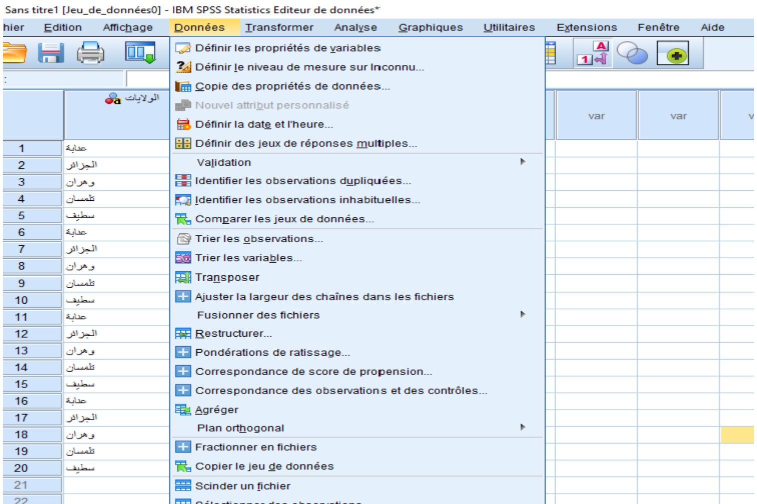
	الولايات	القطاعات	عددالمؤسسات	var
1	عنابة	الصناعة	1912	
2	الجزائر	الصناعة	2452	
3	وهران	الصناعة	5212	
4	تلمسان	الصناعة	9071	
5	سطيف	الصناعة	2021	
6	عنابة	البناء	258	
7	الجزائر	البناء	644	
8	وهران	البناء	185	
9	تلمسان	البناء	171	
10	سطيف	البناء	2230	
11	عنابة	التجارة	10519	
12	الجزائر	التجارة	16175	
13	وهران	التجارة	51618	
14	تلمسان	التجارة	17549	
15	سطيف	التجارة	15888	
16	عنابة	الخدمات	7210	
17	الجزائر	الخدمات	62734	
18	وهران	الخدمات	16297	
19	تلمسان	الخدمات	9472	
20	سطيف	الخدمات	15996	

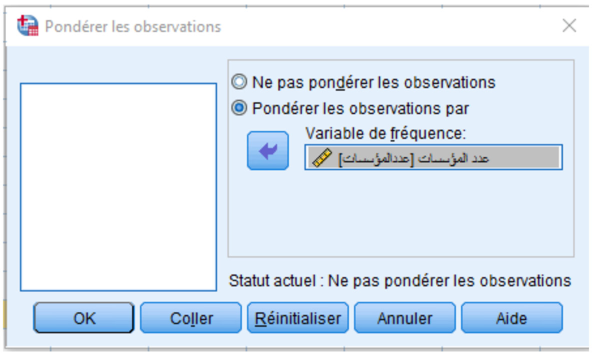
لإنشاء جدول التقاطع الثنائي (جدول التو افق) في برنامج SPSS انطلاقًا من بيانات مُجمَّعة، يستوجب الأمر في مرحلة أولى إجراء عملية الترجيح (Pondération)، وذلك بتحديد متغير التكرارات باعتباره متغير الوزن، حتى يتعامل البرنامج مع كل سطر على أساس عدد المشاهدات الحقيقية المقابلة له لا على أساس مشاهدة واحدة.

الترجيح(Pondération)

Données → Pondérer les observations → Pondérer les observations par → choisir effectif

هذه الخطوة تخبر SPSS بأن كل سطر يمثل n_{ij} مشاهدة وليس مشاهدة واحدة.



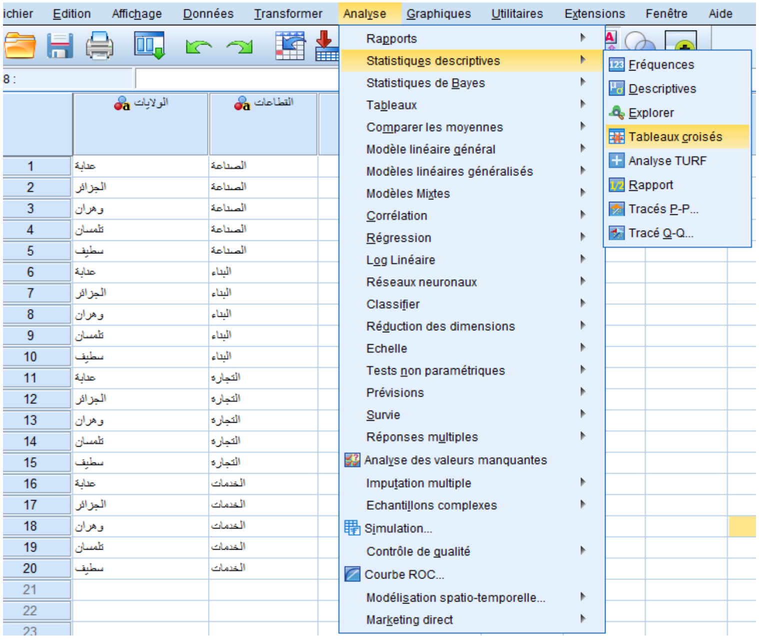


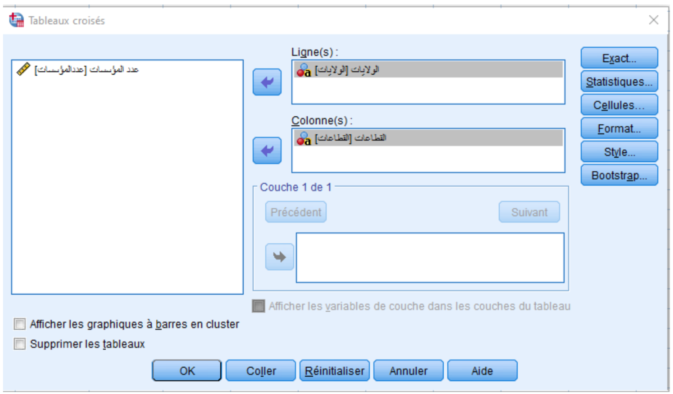
الجدول المزدوج :

Analyse → Statistiques descriptives → Tableaux croisés

الاسطر: الولايات

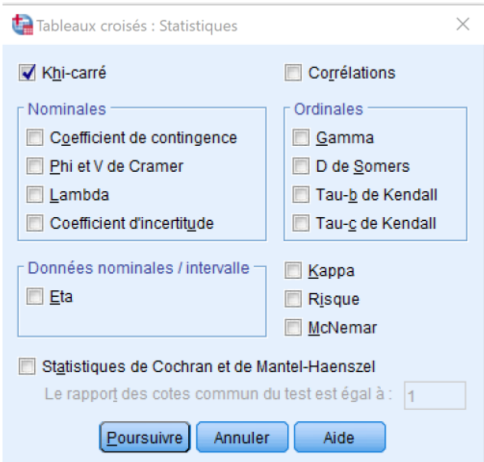
الاعمدة : القطاعات

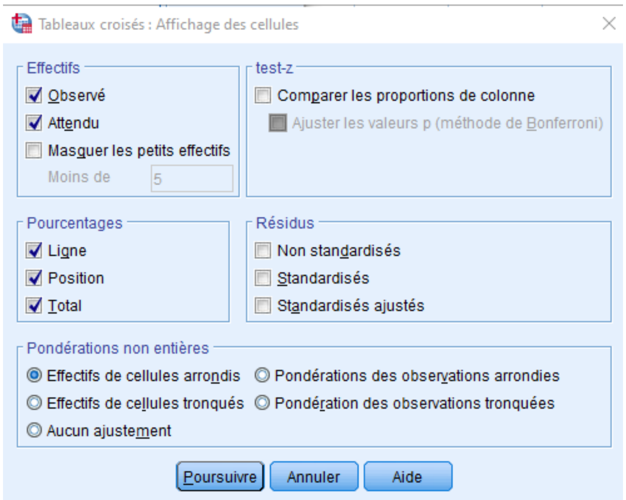




الإحصائية: علم على χ2

الخلايا : علم : Observé / Théorique / Pourcentages en ligne / Pourcentages en





النتائج تكون كالتالي :

Récapitulatif de traitement des observations

Observations					
Valide		Manquant		Total	
N	Pourcentage	N	Pourcentage	N	Pourcentage

القطاعات * الولايات	247614	100,0%	0	0,0%	247614	100,0%

يبين الجدول عدد المشاهدات الكلية ، مع التأكيد من عدم وجود قيم مفقودة.

Tableau croisé القطاعات * الولايات

			القطاعات				
			الصناعة	البناء	التجارة	الخدمات	Total
الولايات	عنابة	Effectif	1912	258	10519	7210	19899
		Effectif théorique	1660,9	280,3	8980,5	8977,3	19899,0
		% dans الولايات	9,6%	1,3%	52,9%	36,2%	100,0%
		% dans القطاعات	9,3%	7,4%	9,4%	6,5%	8,0%
		% du total	0,8%	0,1%	4,2%	2,9%	8,0%
	الجزائر	Effectif	2452	644	16175	62734	82005
		Effectif théorique	6844,8	1155,2	37009,1	36995,9	82005,0
		% dans الولايات	3,0%	0,8%	19,7%	76,5%	100,0%
		% dans القطاعات	11,9%	18,5%	14,5%	56,2%	33,1%
		% du total	1,0%	0,3%	6,5%	25,3%	33,1%
	وهران	Effectif	5212	185	51618	16297	73312
		Effectif théorique	6119,3	1032,7	33085,9	33074,1	73312,0
		% dans الولايات	7,1%	0,3%	70,4%	22,2%	100,0%
		% dans القطاعات	25,2%	5,3%	46,2%	14,6%	29,6%
		% du total	2,1%	0,1%	20,8%	6,6%	29,6%
	تلمسان	Effectif	9071	171	17549	9472	36263
		Effectif théorique	3026,8	510,8	16365,6	16359,8	36263,0
		% dans الولايات	25,0%	0,5%	48,4%	26,1%	100,0%
		% dans القطاعات	43,9%	4,9%	15,7%	8,5%	14,6%
		% du total	3,7%	0,1%	7,1%	3,8%	14,6%
	سطيف	Effectif	2021	2230	15888	15996	36135
		Effectif théorique	3016,1	509,0	16307,8	16302,0	36135,0
		% dans الولايات	5,6%	6,2%	44,0%	44,3%	100,0%
		% dans القطاعات	9,8%	63,9%	14,2%	14,3%	14,6%
		% du total	0,8%	0,9%	6,4%	6,5%	14,6%
Total		Effectif	20668	3488	111749	111709	247614
		Effectif théorique	20668,0	3488,0	111749,0	111709,0	247614,0
		% dans الولايات	8,3%	1,4%	45,1%	45,1%	100,0%
		% dans القطاعات	100,0%	100,0%	100,0%	100,0%	100,0%
		% du total	8,3%	1,4%	45,1%	45,1%	100,0%

يبين الجدول التوافقي العلاقة بين الولايات والقطاعات الاقتصادية، حيث يظهر توزيع المؤسسات أو الأفراد حسب النشاط الاقتصادي داخل كل ولاية، مع مقارنة التكرارات المشاهدة بالتكرارات النظرية في حالة الاستقلال.

نلاحظ أن قطاعي التجارة والخدمات يمثلان النسبة الأكبر من إجمالي الأنشطة، إذ يشكل كل منهما حوالي 45.1% من المجموع الكلي، بينما تبقى الصناعة بنسبة 8.3% والبناء بنسبة 1.4% فقط.

تتميز ولاية الجزائر بسيطرة واضحة لقطاع الخدمات، الذي يمثل 76.5% من أنشطة الولاية، كما تستحوذ وحدها على 56.2% من إجمالي الخدمات في الجدول، وهو ما يعكس الطابع الإداري والخدماتي للعاصمة. بالمقابل، يظهر قطاع التجارة بنسبة أقل نسبيًا.(19.7%)

أما ولاية وهران فتتميز بتركيز مرتفع جدًا في قطاع التجارة، حيث يمثل 70.4% من أنشطة الولاية، كما تستحوذ على 46.2% من إجمالي التجارة، ما يدل على مكانتها التجارية والاقتصادية المهمة.

ولاية تلمسان تُظهر حضورًا قويًا لقطاع الصناعة، إذ يمثل 25% من أنشطة الولاية، كما تضم 43.9% من إجمالي النشاط الصناعي، وهو مستوى أعلى بكثير من المتوقع نظريًا، مما يشير إلى تخصص صناعي واضح.

في سطيف يبرز قطاع البناء بشكل لافت، حيث تمثل الولاية 63.9% من إجمالي نشاط البناء، وهي نسبة مرتفعة جدًا مقارنة بباقي الولايات، إضافة إلى توازن نسبي بين التجارة والخدمات داخل الولاية.

أما عنابة فتتوزع أنشطتها أساسًا بين التجارة (52.9%) والخدمات (36.2%)، مع مساهمة صناعية متوسطة مقارنة ببقية الولايات.

ومن خلال مقارنة التكرارات المشاهدة بالتكرارات النظرية، نلاحظ وجود فروق معتبرة في عدة خلايا، خاصة في:

- الخدمات في الجزائر،
- التجارة في وهران،
- الصناعة في تلمسان،
- البناء في سطيف،

مما يدل على وجود علاقة ارتباط بين الولاية ونوع القطاع الاقتصادي، أي أن توزيع القطاعات ليس مستقلاً عن الولايات، بل إن كل ولاية تتميز بتخصص اقتصادي معين.

➢ **اختبار العلاقة بين المتغيران :**

Tests du khi-carré

	Valeur	ddl	Signification asymptotique (bilatérale)
khi-carré de Pearson	74496,518[a]	12	,000
Rapport de vraisemblance	70065,540	12	,000
N d'observations valides	247614		

a. 0 cellules (0,0%) ont un effectif théorique inférieur à 5. L'effectif théorique minimum est de 280,31.

تشير نتائج اختبار كاي مربع (Khi-carré) إلى وجود علاقة ذات دلالة إحصائية قوية بين الولايات والقطاعات الاقتصادية.

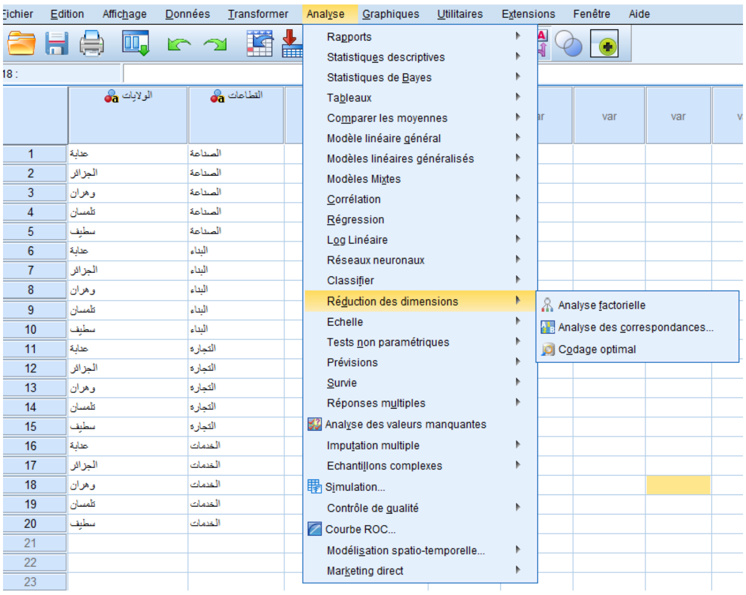
فقد بلغت قيمة اختبار بيرسون لكاي مربع 74496.518 بدرجات حرية تساوي 12، مع مستوى دلالة إحصائية أقل من 0.001 (siG=0.000)، وهو ما يدفع إلى رفض فرضية العدم التي تنص على استقلالية المتغيرين. وبالتالي، نستنتج أن توزيع القطاعات الاقتصادية يختلف باختلاف الولايات.

كما تؤكد قيمة نسبة الإمكان الأعظم (Rapport de vraisemblance) البالغة 70065.540 نفس النتيجة، مما يعزز وجود ارتباط معنوي بين المتغيرين.

إضافة إلى ذلك، فإن شروط تطبيق اختبار كاي مربع متحققة، إذ لا توجد أي خلية ذات تكرار نظري أقل من 5، وكان أصغر تكرار نظري يساوي 280.31، مما يجعل نتائج الاختبار موثوقة إحصائيًا.

وعليه، يمكن القول إن لكل ولاية خصوصية اقتصادية وقطاعية مميزة، وهو ما ظهر أيضًا في الجدول التقاطعي من خلال تمركز بعض القطاعات في ولايات معينة مثل:

- الخدمات في الجزائر،
- التجارة في وهران،
- الصناعة في تلمسان،
- البناء في سطيف.



➢ إجراء التحليل العاملي التقابلي (AFC – Analyse Factorielle des Correspondances)باستخدام برنامج IBM SPSS Statistics v25

من الشريط العلوي نختار:

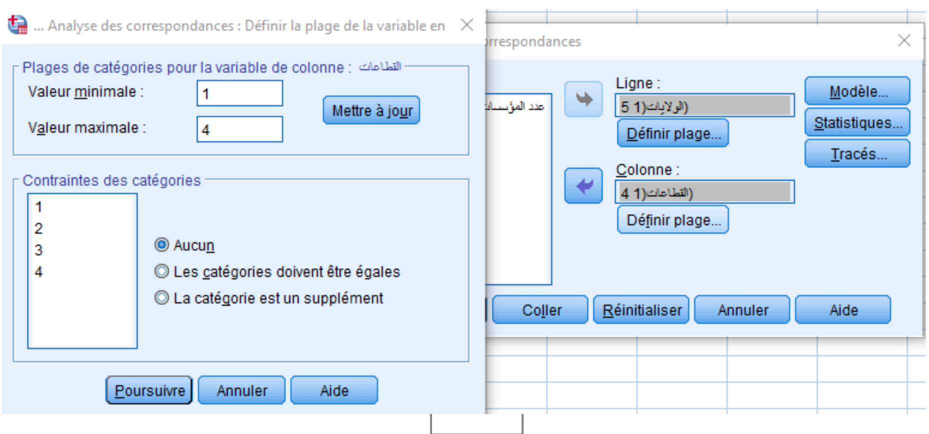
- Analyse
- ثم Réduction de dimension
- ثم Analyse des correspondances.

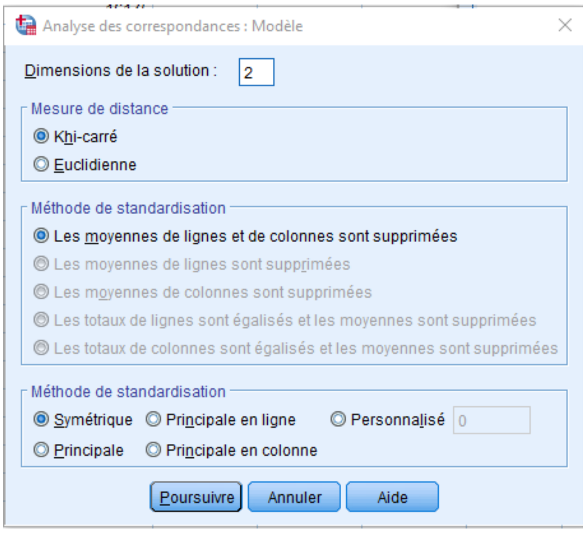
تحديد نطاق البيانات: Définir la plage

بالنسبة للولايات :اقل قيمة هي 1، واعلى قيمة هي 5

بالنسبة للقطاعات:اقل قيمة هي 1، واعلى قيمة هي4

ثم نضغط على تحديث mettre a jour ثم مواصلة : poursuivre





النموذج : Modèle d'analyse des correspondances

هذه النافذة تستخدم لضبط معلمات **تحليل العاملي التو افقي** في البرامج الإحصائية، حيث تتيح للمستخدم التحكم في:

1. **عدد الأبعاد** المستخرجة لتمثيل البيانات بشكل أمثل.
2. **نوع المسافة** المستخدمة لحساب التشابه بين الفئات (كاي تربيع أو إقليدية).
3. **طريقة التقييس** لتحديد كيفية مركزية البيانات (صفوف، أعمدة، أو كلهما).
4. **طريقة التطبيع** لتحديد كيفية عرض النتائج في المخططات الثنائية، حسب الهدف من التحليل: مقارنة المتغيرين معًا، أو التركيز على أحدهما.

تتيح لك هذه النافذة تحديد إعدادات تحليل العاملي (afc) ، وتشمل:

1. أبعاد الحل(Dimensions of the Solution)

- الوظيفة :تحديد عدد الأبعاد المراد استخراجها في التحليل.
- اختر عدد الأبعاد الذي يشرح أقصى قدر من التباين في البيانات.
- الحد الأقصى للأبعاد :يحسب كأصغر قيمة بين:
  - عدد فئات الصفوف النشطة - عدد فئات الصفوف المقيدة + عدد مجموعات فئات الصفوف المقيدة
  - عدد فئات الأعمدة النشطة - عدد فئات الأعمدة المقيدة + عدد مجموعات فئات الأعمدة المقيدة

2. مقياس المسافة(Distance Measure)

الاستخدام	الوصف	الخيار
مطلوب لتحليل العاملي القياسي	مسافة مقطع مرجحة بكتلة الصفوف/الأعمدة	كاي تربيع (Chi-square)
بديل عند عدم استخدام كاي تربيع	الجذر التربيعي لمجموع الفروق المربعة بين الأزواج	إقليدية (Euclidean)

3. طريقة التقييس(Standardization Method)

الوصف	الخيار
مركزية كلا الاتجاهين (مطلوب للتحليل القياسي)	إزالة متوسطات الصفوف والأعمدة

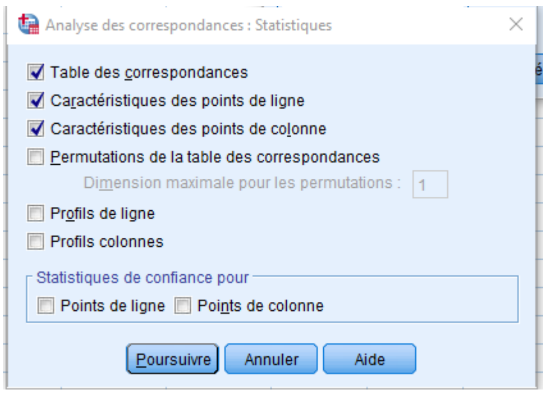
الوصف	الخيار
مركزية الصفوف دون الأعمدة	إزالة متوسطات الصفوف فقط
مركزية الأعمدة دون الصفوف	إزالة متوسطات الأعمدة فقط
توحيد هوامش الصفوف قبل المركزية	تساوي مجاميع الصفوف + إزالة المتوسطات
توحيد هوامش الأعمدة قبل المركزية	تساوي مجاميع الأعمدة + إزالة المتوسطات

4. طريقة التقييس النهائية(Normalization Method)

متى تُستخدم؟	الوصف	الخيار
لمقارنة أوجه التشابه/الاختلاف بين فئات المتغيرين معًا	درجات الصفوف والأعمدة تمثل متوسطات مرجحة مقسومة على القيم المفردة	متماثل (Symmetrical)
لفحص الاختلافات داخل متغيرواحد أو كليهما	المسافات بين النقاط تقريبًا لمسافات جدول التطابق	رئيسي (Principal)
للتركيز على اختلافات فئات متغير الصفوف	مسافات نقاط الصفوف تعكس جدول التطابق	رئيسي للصفوف (Row Principal)
للتركيز على اختلافات فئة متغير الأعمدة	مسافات نقاط الأعمدة تعكس جدول التطابق	رئيسي للأعمدة (Column Principal)
لإنشاء مخططات ثنائية مخصصة حسب الحاجة	قيمة بين -1 و1: • رئيسي للأعمدة = -1 • متماثل = 0 • رئيسي للصفوف = 1	مخصص (Custom)

ملاحظة : لتحليل عاملي توافقي قياسي، استخدم :كاي تربيع كمسافة + إزالة متوسطات الصفوف والأعمدة +التقييس المتماثل.

➢ إحصائيات تحلیل التطابق (Correspondence Analysis Statistics)



تستخدم نافذة "الإحصائيات" في تحليل التطابق لاختيار المخرجات الرقمية التي تساعد على:

1. فهم هيكل البيانات عبر جدول التطابق وملفات الصفوف/الأعمدة.
2. تقييم أهمية الفئات من خلال الكتلة، القصور الذاتي، والمساهمات.
3. تحسين قراءة المخططات عبر الجداول المُبدَّلة حسب الدرجات.
4. قياس دقة التقديرات عبر إحصائيات الثقة (انحراف معياري وارتباطات).

تتيح لك هذه النافذة تحديد المخرجات الرقمية التي ترغب في الحصول عليها من تحليل التطابق:

1. جدول التطابق(Correspondence Table) : جدول توافقي (مصفوفة) للمتغيرات المدخلة، يشمل المجاميع الهامشية للصفوف والأعمدة.

2. خصائص نقاط الصفوف(Row Points Characteristics) :

لكل فئة من فئات الصفوف، يعرض:

العنصر	الوصف
الدرجات (Scores)	إحداثيات النقطة في الأبعاد المستخرجة
الكتلة (Mass)	الوزن النسبي للفئة في الجدول
القصور الذاتي (Inertia)	مقدار التباين الذي تمثله الفئة
مساهمة النقطة في بُعد	مدى مساهمة الفئة في تشكيل بُعد معين
مساهمة البُعد في النقطة	مدى تفسير الأبعاد لموقع الفئة

3. خصائص نقاط الأعمدة(Column Points Characteristics) : نفس المؤشرات السابقة، لكن لفئات الأعمدة.

4. ملفات الصفوف(Row Profiles) : يوضح التوزيع النسبي لكل فئة صف عبر فئات الأعمدة (احتمالات شرطية).

5. ملفات الأعمدة(Column Profiles) : يوضح التوزيع النسبي لكل فئة عمود عبر فئات الصفوف.

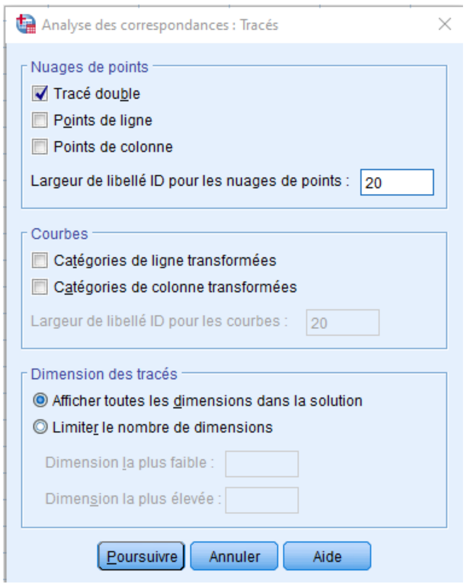
6. تبديلات جدول التطابق(Permutations of the Table) :

- إعادة ترتيب الجدول بحيث تظهر الصفوف والأعمدة بترتيب تصاعدي حسب درجات البُعد الأول.
- خيار إضافي :تحديد الحد الأقصى للأبعاد التي تريد إنشاء جداول مُبدَّلة لها.

7. إحصائيات الثقة لنقاط الصفوف والأعمدة:

النوع	المحتوى
للصفوف	الانحراف المعياري ومعاملات الارتباط لنقاط الصفوف (غير الإضافية)
للأعمدة	الانحراف المعياري ومعاملات الارتباط لنقاط الأعمدة (غير الإضافية)

ملاحظة :نبدأ دائما بطلب "*خصائص النقاط*" و*"مقطع الصفوف/الأعمدة"* لفهم البيانات قبل الانتقال إلى خيارات متقدمة مثل "*التبديلات*" أو"*إحصائيات الثقة*".



➢ المخططات في تحليل العاملي التو افقي (Correspondence Analysis Plots)

تستخدم نافذة "المخططات" لتصميم المخرجات الرسومية لتحليل التطابق، وتشمل ثلاث وظائف رئيسية:

1. المخططات الانتشارية: لعرض العلاقات بين النقاط في فضاء الأبعاد (ثنائية أو منفصلة للصفوف/الأعمدة).
2. المخططات الخطية: لمقارنة القيم الأصلية للفئات مع درجاتها المستخرجة في التحليل.
3. التحكم في الأبعاد: لعرض جميع الأبعاد أو اختيار نطاق محدد لتجنب الازدحام البصري.

تتيح لك هذه النافذة تحديد أنواع الرسوم البيانية التي ترغب في إنشائها لتحليل التطابق:

1. المخططات الانتشارية(Scatterplots / Nuages de points) : تولد مصفوفة من جميع المخططات التي تعرض الأبعاد بشكل أزواج

2. المخططات الخطية(Line Plots / Courbes) : ينشئ مخططا منفصلا لكل بُعد من المتغير المحدد.

3. **أبعاد المخططات(Plot Dimensions)** : التحكم في الأبعاد التي تظهر في المخرجات الرسومية.

ملاحظة :

- استخدم المخطط الثنائي (Biplot) فقط مع التقييس "المتماثل.(Symmetrical) "
- حدّد عدد أحرف التسميات بـ 10-15 حرفًا لتوازن بين الوضوح وعدم الازدحام.
- عند وجود 5<أبعاد، استخدم خيار"تحديد الأبعاد" لعرض الأزواج الأكثر أهمية (مثل: البُعد 1 مقابل 2).

ننقرعلى التالي poursuivre، ثم ok، نتائج التحليل تظهر كمايلي :

1. الجدول التوافقي الأصلي :

Table des correspondances

الولايات	القطاعات الصناعة	البناء	التجارة	الخدمات	Marge active
عنابة	1912	258	10519	7210	19899
الجزائر	2452	644	16175	62734	82005
وهران	5212	185	51618	16297	73312
تلمسان	9071	171	17549	9472	36263
سطيف	2021	2230	15888	15996	36135
Marge active	20668	3488	111749	111709	247614

ملخص النموذج(Récapitulatif) :

Récapitulatif

Dimension	Valeur singulière	Inertie	Khi-deux	Sig.	Proportion d'inertie: Représentation	Proportion d'inertie: Cumulé	Valeur singulière de confiance: Ecart type	Valeur singulière de confiance: Corrélation 2
1	,473	,223			,742	,742	,002	,090
2	,224	,050			,167	,910	,003	
3	,165	,027			,090	1,000		
Total		,301	74496,518	,000[a]	1,000	1,000		

a. 12 degrés de liberté

- **Khi-deux** = 74,496.518 مهم جداً، ($p < 0.001$)
- **البعدان الأولان** يفسران **91%** من التباين الكلي
- هذا يعني أن التمثيل في بعدين **كافي ودقيق**

تحليل نقاط الصفوف (الدول)

Présentation des points de ligne[a]

الولايات	Masse	Score de la dimension: 1	Score de la dimension: 2	Inertie	Contribution: Du point vers l'inertie de la dimension: 1	Contribution: Du point vers l'inertie de la dimension: 2	Contribution: De la dimension vers l'inertie du point: 1	Contribution: De la dimension vers l'inertie du point: 2	Contribution: De la dimension vers l'inertie du point: Total
عنابة	,080	-,262	-,028	,003	,012	,000	,994	,005	1,000
الجزائر	,331	,910	,050	,132	,581	,004	,983	,001	,984
وهران	,296	-,681	-,401	,080	,291	,213	,816	,134	,950
تلمسان	,146	-,609	1,044	,062	,115	,712	,416	,581	,996
سطيف	,146	,072	-,332	,025	,002	,072	,014	,145	,159
Total actif	1,000			,301	1,000	1,000			

a. Normalisation symétrique

- الجزائر :تساهم بـ 58.1%في تكوين البعد الأول)الأهم(
- تلمسان :تساهم بـ 71.2%في تكوين البعد الثاني
- وهران :تساهم في كلا البعدين بشكل متوازن

تحليل نقاط الأعمدة (القطاعات)

Présentation des points de colonne[a]

القطاعات	Masse	Score de la dimension: 1	Score de la dimension: 2	Inertie	Contribution: Du point vers l'inertie de la dimension: 1	Contribution: Du point vers l'inertie de la dimension: 2	Contribution: De la dimension vers l'inertie du point: 1	Contribution: De la dimension vers l'inertie du point: 2	Contribution: De la dimension vers l'inertie du point: Total
الصناعة	,083	-,737	1,462	,062	,096	,795	,345	,644	,989
البناء	,014	,273	-,780	,028	,002	,038	,018	,068	,086
التجارة	,451	-,620	-,285	,091	,367	,163	,904	,091	,995
الخدمات	,451	,748	,039	,120	,535	,003	,996	,001	,997
Total actif	1,000			,301	1,000	1,000			

a. Normalisation symétrique

- الخدمات : تساهم بـ 53.5%في البعد الأول (القطاع الأكبر كتلة: 45.1%)

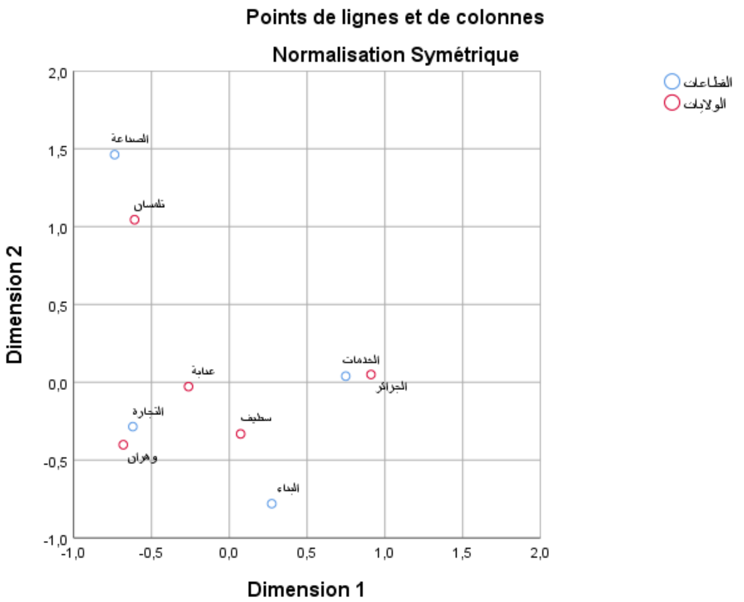
- ➢ الصناعة: تساهم بـ 79.5%في البعد الثاني رغم كتلتها الصغيرة(8.3%)
- ➢ التجارة: تساهم بشكل كبير في البعد الأول(36.7%)

Points de ligne de confiance

الولايات	Ecart type de la dimension 1	Ecart type de la dimension 2	Corrélation 1-2
عنابة	,001	,001	-,135
الجزائر	,003	,006	-,079
وهران	,004	,009	-,253
تلمسان	,009	,007	,190
سطيف	,009	,024	,073

Points de colonne de confiance

القطاعات	Ecart type de la dimension 1	Ecart type de la dimension 2	Corrélation 1-2
الصناعة	,012	,011	,129
البناء	,021	,082	-,016
التجارة	,003	,004	-,496
الخدمات	,002	,003	-,115



➢ التفسير العام للأبعاد

البعد الأول (74.2% من التباين):

- ➢ الاتجاه الموجب: الجزائر (+0.910) ⟵⟶ الخدمات(+0.748)
- ➢ الاتجاه السالب: وهران (-0.681)، تلمسان (-0.609) ⟵⟶ التجارة(-0.620)

التفسير: يميز هذا البعد بين:

- ➢ ولايات تركز على الخدمات (الجزائر، سطيف)
- ➢ ولايات تركز على التجارة (وهران، تلمسان)

البعد الثاني (16.7% من التباين):

- ➢ الاتجاه الموجب: تلمسان (+1.044) ⟵⟶ الصناعة(+1.462)
- ➢ الاتجاه السالب: وهران (-0.401) ⟵⟶ البناء(-0.780)

التفسير :يميز هذا البعد بين:

- ولايات صناعية (تلمسان)
- ولايات تركز على البناء (وهران)

الخلاصة:

1. النموذج يفسر 91% من التباين في بعدين فقط
2. العلاقة بين ولايات والقطاعات مهمة إحصائياً ($p < 0.001$)

الاستنتاجات:

- الجزائر تتميز بقطاع الخدمات
- تلمسان تتميز بقطاع الصناعة
- وهران تتميز بقطاع التجارة
- الخدمات والتجارة هما القطاعان المهيمنان (45.1% لكل منهما)

توصيات:

- عنابة تحتاج إلى تنويع اقتصادي (مساهمته ضعيفة في كلا البعدين)
- سطيف متوازنة نسبياً بين القطاعات

تمارين :

تمرين تطبيقي :

تعطي العينة التالية توزيع 10000 طالب من حيث نوع الدراسة والفئة الاجتماعية-المهنية للأب:

الفئة الاجتماعية-المهنية للأب	الحقوق	العلوم الاقتصادية	الآداب	العلوم	الطب	الصيدلة	متعدد التخصصات	معاهد التكنولوجيا IUT
مستغل فلاحي	80	36	134	99	65	28	11	58
أجير	6	2	15	6	4	1	1	4
ربّ عمل	168	74	312	137	208	53	21	62
مهنة حرة	470	191	806	400	876	164	45	79
إطار متوسط	236	99	493	264	281	56	36	87
موظف	145	52	281	133	135	30	20	54
عامل	166	64	401	193	127	23	28	129

معاهد التكنولوجيا IUT	متعدد التخصصات	الصيدلة	الطب	العلوم	الآداب	العلوم الاقتصادية	الحقوق	الفئة الاجتماعية-المهنية للأب
8	2	2	8	11	27	6	16	خدمات
90	42	47	301	247	624	115	305	أخرى

المطلوب: اجراء تحليل العاملي توافقي باستخدام برنامج SPSS:

تمرين الاعمال تطبيقية:

أُجري استبيان حول تقييم 5 منتجات من قبل 4 فئات عمرية مختلفة .تم تسجيل عدد الأشخاص الذين اختاروا كل تقييم حسب الفئة والمنتج.

	منتجA	منتجB	منتجC	منتجD	منتجE
18-25سنة	20	35	15	25	5
26-40سنة	30	20	25	15	10
41-60سنة	15	25	30	20	10
60+سنة	5	10	20	30	35

المطلوب:

ما نوع البيانات الموجودة في هذا الجدول؟ هل هي مناسبة لتحليلAFC ؟

احسب الجداول التالية:

-جدول التكرارات النسبية.(fréquences relatives)

-مجموعات الصفوف والأعمدة.(profils ligne et colonne)

استخدم برنامجًا مثل XLSTAT أو SPSS لإجراء تحليل AFC لهذا الجدول.

مثل النتائج في مستوى ثنائي الأبعاد، وفسّر العلاقة بين الفئات العمرية والمنتجات.

ماذا تستنتج من التمثيل البياني؟ هل هناك فئة تفضل منتجًا معينًا بوضوح؟

قائمة المراجع:

1. الحميد، إ. م. (2015). تحليل البيانات متعددة المتغيرات باستخدام الحاسوب. جامعة القصيم.
2. Lay, D. C., Lay, S. R., & McDonald, J. J. (2021). Linear Algebra and Its Applications (6th ed.). Pearson.
3. Strang, G. (2016). Introduction to Linear Algebra (5th ed.). Wellesley-Cambridge Press.
4. Husson, F., Lê, S., & Pagès, J. (2017). Exploratory Multivariate Analysis by Example Using R (2nd ed.). CRC Press.
5. Jolliffe, I. T. (2002). Principal Component Analysis (2nd ed.). Springer.
6. Pagès, J. (2014). Multiple Factor Analysis by Example Using R. CRC Press.
7. Wickham, H., & Grolemund, G. (2017). R for Data Science. O'Reilly Media.
8. Lê, S., Josse, J., & Husson, F. (2008). FactoMineR: An R Package for Multivariate Analysis. Journal of Statistical Software, 25(1), 1–18.
9. McKinney, W. (2017). Python for Data Analysis (2nd ed.). O'Reilly Media.
10. Jean-Marc Lasgouttes, Variables qualitatives : analyse des correspondances, http://ana-donnees.lasgouttes.net/
11. IBM SPSS Statistics , https://www.ibm.com/docs/fr/spss-statistics/cd?topic=categories-correspondence-analysis.
12. Jean-Marc Lasgouttes, Cours d'analyse de données, INSA Rouen Normandie, Département Génie Mathématique, 2029/2024, https://who.paris.inria.fr/Jean-Marc.Lasgouttes/ana-donnees/
13. Jean Stafford, Paul Bodson, L'analyse multivariée avec SPSS, Presses de l'Université du Québec, 2006
14. Ricco Rakotomalala , Pratique des Méthodes Factorielles avec Python, Version 1.0, Université Lumière Lyon 2,2020

الملاحق

نبذة عن الاعلام المذكورين في المطبوعة[6]:

1. سبيرمان، تشارلز إدوارد (1863-1945) SPEARMAN, Charles Edward

عالم نفس إنجليزي وتلميذ للعالم فونت Wundt ، اشتهر بأعماله حول الذكاء، حيث افترض وجود عامل عام للذكاء (g factor) يكمن وراء جميع الأنشطة الفكرية، كما قدم مساهمات مهمة في التحليل العاملي والإحصاء، خاصة في مجال الارتباطات، بدأ سبيرمان أبحاثه في وقت متأخر، إذ بدأ الدكتوراه في سن 34 عامًا.

يُعتبر أحد رواد النظرية الكلاسيكية للاختبارات (Boake, 2002) وساهم في تطوير مؤشرات مثل الثبات(Fidélité) ، ولا يزال من أبرز الباحثين في مجال قياس الذكاء، إلى جانب كاتل Cattell ، بينيه Binet ، وجالتون Galton .

2. ثورستون، لويس ليون (1887-1955) THURSTONE Louis Leon

عالم نفس بارز في **المدرسة الأمريكية لعلم النفس**، بدأ مسيرته المهنية كمهندس كهربائي، حيث اخترع **جهاز عرض سينمائي** وعمل لفترة قصيرة كمساعد لتوماس إديسون. قادته أبحاثه لاحقًا إلى دراسة **القياس الحسي** وعلم النفس، فحصل على **درجة الدكتوراه** من جامعة شيكاغو عام 1917، وكرّس بقية حياته لدراسة علم النفس.

عارض **نظرية سبيرمان** حول الذكاء العام، وطور **أساليب جديدة للتحليل العاملي**، حيث حدد **القدرات العقلية الأولية** مثل :**الفهم اللفظي، الطلاقة اللفظية، الاستدلال، وغيرها** .يُعرف بشكل أساسي بأعماله في **التحليل العاملي في دراسة الذكاء**، لكنه أسهم أيضًا في **القياس النفسي** وعلم النفس الاجتماعي.

3. بِنزِكري، جان بول (1932-2019) Jean-Paul Benzécri

إحصائي فرنسي من مواليد وهران-الجزائر، وهو خريج المدرسة العليا للأساتذة، أجرى أطروحة دكتوراه عام 1955 تحت إشراف كارتون Carton ، وكانت في مجال التوبولوجيا. يُعد مؤسس المدرسة الفرنسية لتحليل البيانات، واشتهر بشكل خاص بتطوير التحليل العاملي التناظري(AFC) .

بدأ مسيرته الأكاديمية كباحث وأستاذ في جامعة رين، ثم أصبح أستاذًا في كلية العلوم بباريس، وبعدها في جامعة بيير وماري كوري. في شبابه، تواصل مع «مجموعة القياس متعدد الأبعاد-Multi) » « Dimensional Scaling group)في الولايات المتحدة، لكن أبحاثه لم تلقَ ترحيبًا في العالم الأنجلوساكسوني، ربما نتيجة زيارته عام 1965 لمختبرات Bell Telephone Laboratories . المعروفة حاليا

[6] للتعمق اكثر يرجى الاطلاع على الموقع : https://www.intelltheory.com/map.shtml

باسم Bell Labs ، وهي واحدة من أهم مراكز البحث والتطوير في العالم، وقد لعبت دورًا رئيسيًا في تطوير العديد من الابتكارات التكنولوجية والعلمية، وذلك بسبب الاختلاف المنهجي بين المدرسة الفرنسية لتحليل البيانات والمناهج الأمريكية في القياس متعدد الأبعاد (MDS)

نتيجة لذلك، تم تجاهل بعض من تطوراته في التحليل العاملي التناظري (AFC)، مما أدى إلى عزلته العلمية، لكنه استمر في البحث والتأمل، مما أثمر عن تأليف عمل هام حول تاريخ وتحليل البيانات، نُشر عام 1982.

4. هارولد هوتيلينغ(1895-1973) Harold Hotelling

عالم رياضيات وإحصائي أمريكي، قدم مساهمات بارزة في الإحصاء، الاقتصاد، والرياضيات التطبيقية، اشتهر بتطوير تحليل المكونات الرئيسية (ACP – Principal Component Analysis) عام 1933، وهي تقنية تُستخدم لتقليل أبعاد البيانات الكبيرة مع الحفاظ على أكبر قدر ممكن من المعلومات.

كان له تأثير كبير في نظرية القرار الإحصائي وتحليل التباين التمييزي، كما ساهم في الاقتصاد من خلال نظرية ندرة الموارد، والتي أثرت على السياسات البيئية والاقتصادية.

شغل منصب رئيس الجمعية الإحصائية الأمريكية وكان له دور في تطوير برامج الإحصاء في الجامعات الأمريكية، ولا تزال أفكاره تُستخدم على نطاق واسع في البيانات الضخمة، الذكاء الاصطناعي، والاقتصاد القياسي.

5. كارل بيرسون(1857-1936) Karl Pearson

عالم إحصاء ورياضي بريطاني، يُعتبر أحد مؤسسي علم الإحصاء الحديث، طوّر معامل الارتباط لبيرسون(Pearson Correlation Coefficient) ، الذي يُستخدم لقياس العلاقة بين متغيرين.

كان من رواد الإحصاء الاستدلالي وأسهم في وضع أسس تحليل الانحدار والاحتمالات، حيث أسس أول قسم للإحصاء في جامعة لندن وأطلق مجلة Biometrika التي لعبت دورًا رئيسيًا في تطوير الإحصاء الحيوي.

تأثر بأفكار فرانسيس جالتون وعمل على تطوير مفهوم الانحراف المعياري والتوزيع الطبيعي. تُستخدم طرقه الإحصائية اليوم في البحث العلمي، علم البيانات، الذكاء الاصطناعي، والاقتصاد القياسي.

Table entry for p and C is the critical value t^* with probability p lying to its right and probability C lying between $-t^*$ and t^*.

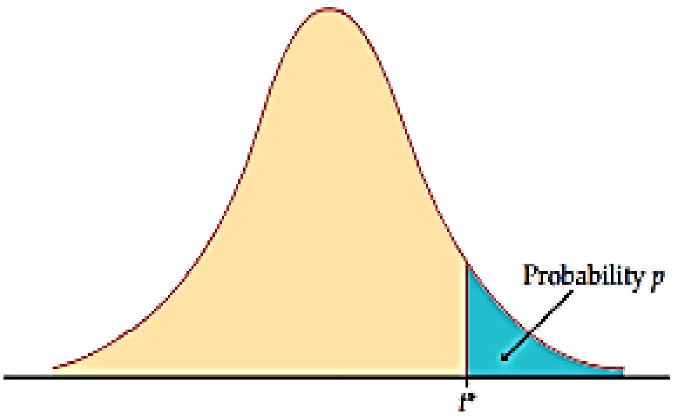
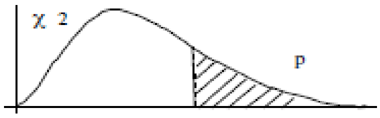


TABLE D

t distribution critical values

	Upper-tail probability p											
df	.25	.20	.15	.10	.05	.025	.02	.01	.005	.0025	.001	.0005
1	1.000	1.376	1.963	3.078	6.314	12.71	15.89	31.82	63.66	127.3	318.3	636.6
2	0.816	1.061	1.386	1.886	2.920	4.303	4.849	6.965	9.925	14.09	22.33	31.60
3	0.765	0.978	1.250	1.638	2.353	3.182	3.482	4.541	5.841	7.453	10.21	12.92
4	0.741	0.941	1.190	1.533	2.132	2.776	2.999	3.747	4.604	5.598	7.173	8.610
5	0.727	0.920	1.156	1.476	2.015	2.571	2.757	3.365	4.032	4.773	5.893	6.869
6	0.718	0.906	1.134	1.440	1.943	2.447	2.612	3.143	3.707	4.317	5.208	5.959
7	0.711	0.896	1.119	1.415	1.895	2.365	2.517	2.998	3.499	4.029	4.785	5.408
8	0.706	0.889	1.108	1.397	1.860	2.306	2.449	2.896	3.355	3.833	4.501	5.041
9	0.703	0.883	1.100	1.383	1.833	2.262	2.398	2.821	3.250	3.690	4.297	4.781
10	0.700	0.879	1.093	1.372	1.812	2.228	2.359	2.764	3.169	3.581	4.144	4.587
11	0.697	0.876	1.088	1.363	1.796	2.201	2.328	2.718	3.106	3.497	4.025	4.437
12	0.695	0.873	1.083	1.356	1.782	2.179	2.303	2.681	3.055	3.428	3.930	4.318
13	0.694	0.870	1.079	1.350	1.771	2.160	2.282	2.650	3.012	3.372	3.852	4.221
14	0.692	0.868	1.076	1.345	1.761	2.145	2.264	2.624	2.977	3.326	3.787	4.140
15	0.691	0.866	1.074	1.341	1.753	2.131	2.249	2.602	2.947	3.286	3.733	4.073
16	0.690	0.865	1.071	1.337	1.746	2.120	2.235	2.583	2.921	3.252	3.686	4.015
17	0.689	0.863	1.069	1.333	1.740	2.110	2.224	2.567	2.898	3.222	3.646	3.965
18	0.688	0.862	1.067	1.330	1.734	2.101	2.214	2.552	2.878	3.197	3.611	3.922
19	0.688	0.861	1.066	1.328	1.729	2.093	2.205	2.539	2.861	3.174	3.579	3.883
20	0.687	0.860	1.064	1.325	1.725	2.086	2.197	2.528	2.845	3.153	3.552	3.850
21	0.686	0.859	1.063	1.323	1.721	2.080	2.189	2.518	2.831	3.135	3.527	3.819
22	0.686	0.858	1.061	1.321	1.717	2.074	2.183	2.508	2.819	3.119	3.505	3.792
23	0.685	0.858	1.060	1.319	1.714	2.069	2.177	2.500	2.807	3.104	3.485	3.768
24	0.685	0.857	1.059	1.318	1.711	2.064	2.172	2.492	2.797	3.091	3.467	3.745
25	0.684	0.856	1.058	1.316	1.708	2.060	2.167	2.485	2.787	3.078	3.450	3.725
26	0.684	0.856	1.058	1.315	1.706	2.056	2.162	2.479	2.779	3.067	3.435	3.707
27	0.684	0.855	1.057	1.314	1.703	2.052	2.158	2.473	2.771	3.057	3.421	3.690
28	0.683	0.855	1.056	1.313	1.701	2.048	2.154	2.467	2.763	3.047	3.408	3.674
29	0.683	0.854	1.055	1.311	1.699	2.045	2.150	2.462	2.756	3.038	3.396	3.659
30	0.683	0.854	1.055	1.310	1.697	2.042	2.147	2.457	2.750	3.030	3.385	3.646
40	0.681	0.851	1.050	1.303	1.684	2.021	2.123	2.423	2.704	2.971	3.307	3.551
50	0.679	0.849	1.047	1.299	1.676	2.009	2.109	2.403	2.678	2.937	3.261	3.496
60	0.679	0.848	1.045	1.296	1.671	2.000	2.099	2.390	2.660	2.915	3.232	3.460
80	0.678	0.846	1.043	1.292	1.664	1.990	2.088	2.374	2.639	2.887	3.195	3.416
100	0.677	0.845	1.042	1.290	1.660	1.984	2.081	2.364	2.626	2.871	3.174	3.390
1000	0.675	0.842	1.037	1.282	1.646	1.962	2.056	2.330	2.581	2.813	3.098	3.300
z^*	0.674	0.841	1.036	1.282	1.645	1.960	2.054	2.326	2.576	2.807	3.091	3.291
	50%	60%	70%	80%	90%	95%	96%	98%	99%	99.5%	99.8%	99.9%
	Confidence level C											

TABLE DU CHI-DEUX : $\chi^2(n)$



n \ p	0.90	0.80	0.70	0.50	0.30	0.20	0.10	0.05	0.02	0.01
1	0,0158	0,0642	0,148	0,455	1,074	1,642	2,706	3,841	5,412	6,635
2	0,211	0,446	0,713	1,386	2,408	3,219	4,605	5,991	7,824	9,210
3	0,584	1,005	1,424	2,366	3,665	4,642	6,251	7,815	9,837	11,341
4	1,064	1,649	2,195	3,357	4,878	5,989	7,779	9,488	11,668	13,277
5	1,610	2,343	3,000	4,351	6,064	7,289	9,236	11,070	13,388	15,086
6	2,204	3,070	3,828	5,348	7,231	8,558	10,645	12,592	15,033	16,812
7	2,833	3,822	4,671	6,346	8,383	9,803	12,017	14,067	16,622	18,475
8	3,490	4,594	5,527	7,344	9,524	11,030	13,362	15,507	18,168	20,090
9	4,168	5,380	6,393	8,343	10,656	12,242	14,684	16,919	19,679	21,666
10	4,865	6,179	7,267	9,342	11,781	13,442	15,987	18,307	21,161	23,209
11	5,578	6,989	8,148	10,341	12,899	14,631	17,275	19,675	22,618	24,725
12	6,304	7,807	9,034	11,340	14,011	15,812	18,549	21,026	24,054	26,217
13	7,042	8,634	9,926	12,340	15,119	16,985	19,812	22,362	25,472	27,688
14	7,790	9,467	10,821	13,339	16,222	18,151	21,064	23,685	26,873	29,141
15	8,547	10,307	11,721	14,339	17,322	19,311	22,307	24,996	28,259	30,578
16	9,312	11,152	12,624	15,338	18,418	20,465	23,542	26,296	29,633	32,000
17	10,085	12,002	13,531	16,338	19,511	21,615	24,769	27,587	30,995	33,409
18	10,865	12,857	14,440	17,338	20,601	22,760	25,989	28,869	32,346	34,805
19	11,651	13,716	15,352	18,338	21,689	23,900	27,204	30,144	33,687	36,191
20	12,443	14,578	16,266	19,337	22,775	25,038	28,412	31,410	35,020	37,566
21	13,240	15,445	17,182	20,337	23,858	26,171	29,615	32,671	36,343	38,932
22	14,041	16,314	18,101	21,337	24,939	27,301	30,813	33,924	37,659	40,289
23	14,848	17,187	19,021	22,337	26,018	28,429	32,007	35,172	38,968	41,638
24	15,659	18,062	19,943	23,337	27,096	29,553	33,196	36,415	40,270	42,980
25	16,473	18,940	20,867	24,337	28,172	30,675	34,382	37,652	41,566	44,314
26	17,292	19,820	21,792	25,336	29,246	31,795	35,563	38,885	42,856	45,642
27	18,114	20,703	22,719	26,336	30,319	32,912	36,741	40,113	44,140	46,963
28	18,939	21,588	23,647	27,336	31,391	34,027	37,916	41,337	45,419	48,278
29	19,768	22,475	24,577	28,336	32,461	35,139	39,087	42,557	46,693	49,588
30	20,599	23,364	25,508	29,336	33,530	36,250	40,256	43,773	47,962	50,892

Pour $n > 30$, on peut admettre que $\sqrt{2\chi^2} - \sqrt{2n-1} = N(0,1)$