جامعة العربي بن مهيدي – أم البواقي-
كلية العلوم الاقتصادية والعلوم التجارية وعلوم التسيير
قسم العلوم الاقتصادية
الاستاذ : بوزيان محمد

السنة الجامعية: 2024 - 2025
المستوى: ثالثة ليسانس
التخصص: اقتصاد كمي
المدة: ساعة ونصف

امتحان الدورة العادية في مقياس تحليل البيانات 2

الجانب النظري : 09 نقاط

1- ما المقصود بطريقة التحليل العاملي للتوفيقات(AFC) ؟ وما نوع الجداول التي يعتمد عليها تحليلAFC ؟
2- ما الهدف من تحويل جدول التكرارات إلى جدول احتمالات أو ما الغرض من استخدام جدول البروفايلات في تحليل AFC ؟
3- لماذا يعتبر AFC امتدادًا لـ ACP في حالة البيانات النوعية؟
4- ما الفائدة من تحليل العطالة فيAFC ؟
5- كيف يتم تفسير موضع فئة في الخريطة العاملية؟
6- في التصنيف الهرمي، ما الفرق بين خوارزمية 'Single Linkage' وخوارزمية'Ward' مع العلم أن خوارزمية 'Single Linkage' تُستخدم لدمج المجموعات أو المفردات الأقرب من حيث أقصر مسافة بين أي عنصر من المجموعة الأولى وأي عنصر من المجموعة الثانية وفق العلاقة التالية : f_0

$$\{B \ni A, y \ni x \mid \min\{d(x,y) = d(A,B)\}$$

7- كيف يمكن تحديد عدد المجموعات المثالي في التصنيف الهرمي؟
8- ما علاقة العطالة الخارجية بالتجانس الداخلي للمجموعات؟
9- متى نُفضل استخدام التصنيف التنازلي بدل التصاعدي؟

الجانب التطبيقي : 11 نقاط

التمرين الاول : (06 نقاط)

لقد سألنا بعض الأشخاص عن عدد الساعات التي يقضونها أسبوعيًا على منصات التواصل الاجتماعي وفي صالة الألعاب الرياضية. وكانت الارقام كما يلي :

	Réseaux sociaux	Salle de sport
Mohamed	2	3
Mouad	5	2
Younes	5	3
Ilyes	1	4
Fatima	4	5

نريد اعداد تحليل باستخدام التصنيف الهرمي التصاعدي . حيث نقوم باعداد تجميع للأفراد في كل مرحلة الى غاية الحصول على مجموعة واحدة .

المطلوب :

1- ماهي الخطوات المتبعة في ذلك ؟ هل تعطي طريقة التصنيف نتيجة واحدة (علل أجابتك)؟
2- قم برسم شجرة التصنيف الموافقة مع خوارزميتك المتبعة ؟ ماهو عدد الفئات الممكن تشكيلها (علل اجابتك)؟
3- احسب كل من العطالة الكلية , مابين المجموعات والعطالة داخل المجموعات ؟

التمرين 2:

Contingency Table

	Non_fum	Occas	Moyen	Gros_fum	Sum
Cadre_exper	4	2	3	2	11
cadre_debut	4	3	7	4	18
Empl_exper	25	10	12	4	51
Empl_debut	18	24	33	13	88
Secretaire	10	6	7	2	25
Sum	61	45	62	25	193

Inertia and Chi-Square Decomposition

Singular Value	Principal Inertia	Chi-Square	Percent	Cumulative Percent	18 36 54 72 90
0.27342	0.07476	14.4285	87.76	87.76	**********
0.10009	0.01002	1.9333	11.76	99.51	***
0.02034	0.00041	0.0798	0.49	100.00	
Total	0.08519	16.4416	100.00		

Degrees of Freedom = 12

Partial Contributions to Inertia for the Row Points

	Dim1	Dim2
Cadre_exper	0.0033	0.2136
cadre_debut	0.0837	0.5512
Empl_exper	0.5120	0.0030
Empl_debut	0.3310	0.1518
Secretaire	0.0701	0.0805

Squared Cosines for the Row Points

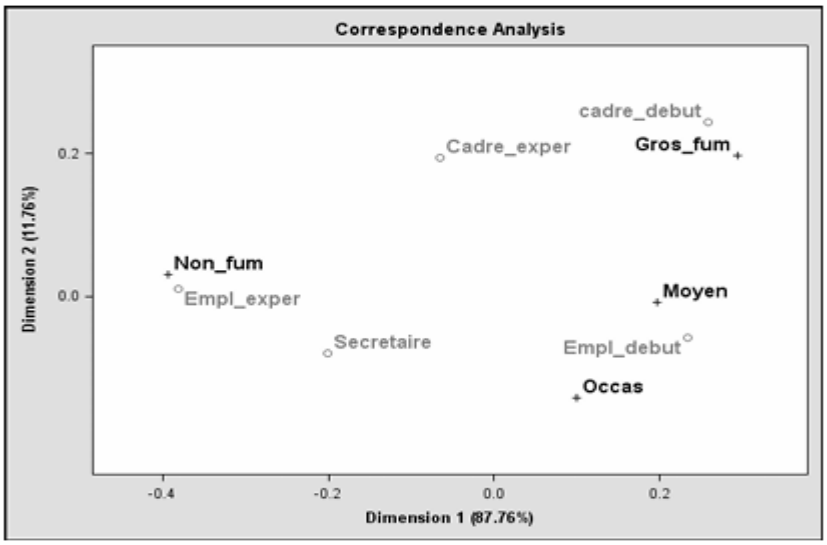
	Dim1	Dim2
Cadre_exper	0.0922	0.8003
cadre_debut	0.5264	0.4647
Empl_exper	0.9990	0.0008
Empl_debut	0.9419	0.0579
Secretaire	0.8653	0.1333

Partial Contributions to Inertia for the Column Points

	Dim1	Dim2
Non_fum	0.6540	0.0293
Occas	0.0308	0.4632
Moyen	0.1656	0.0017
Gros_fum	0.1495	0.5058

Squared Cosines for the Column Points

	Dim1	Dim2
Non_fum	0.9940	0.0060
Occas	0.3267	0.6573
Moyen	0.9818	0.0014
Gros_fum	0.6844	0.3102



المطلوب : قم بتحليل النتائج

جامعة العربي بن مهيدي – أم البواقي-
كلية العلوم الاقتصادية والعلوم التجارية وعلوم التسيير
قسم العلوم الاقتصادية

السنة الجامعية: 2024 - 2025
المستوى ثالثة ليسانس
التخصص اقتصاد كمي
المدة: ساعة ونصف

الاستاذ بوزيان محمد

الإجابة النموذجية لامتحان الدورة العادية في مقياس تحليل البيانات 2

الجانب النظري : 09 نقاط

الجانب النظري :

1- **الإجابة** :هي تقنية تحليل استكشافي متعددة الأبعاد تُستخدم لتمثيل البيانات الوصفية (النوعية) بيانياً من خلال خفض الأبعاد. حيث تقوم على جداول التكرارات التي تربط بين متغيرين وصفيين أو ما يعرف بجدول الاقتران (02 نقاط).

2- **الإجابة :**الهدف هو تحويل القيم المطلقة إلى نسب شرطية (probabilités conditionnelles) توضّح كيفية توزيع الفئات داخل كل متغير، ما يساعد على فهم العلاقات النسبية بين الصفوف والأعمدة، وتمييز التشابهات والاختلافات (01 نقاط)..

3- **الإجابة** :لأن كلاهما يهدف لاختزال الأبعاد، لكن AFC مخصص للبيانات الوصفية. (01 نقاط)

4- **الإجابة** :العطالة تمثل التباين المفسّر، وكلما كانت مرتفعة لمحور ما دل ذلك على أهميته. (01 نقاط).

5- **الإجابة:** بناءً على موقعها على المحاور (F1 وF2)، حيث يشير اتجاهها إلى تقاربها مع فئات أخرى، وبعدها عن المركز يدل على مساهمتها الكبيرة في التباين. (01 نقاط).

6- **الاجابة :** 'Single Linkage' تعتمد على المسافة الدنيا بين عناصر المجموعتين، مما قد يؤدي إلى مجموعات غير متجانسة. أما 'Ward' فتعتمد على تقليل التباين داخل المجموعات (العطالة الداخلية)، مما يعطي مجموعات أكثر تجانسًا. (01 نقاط).

7- **الاجابة :** من خلال تحليل الشجرة (Dendrogramme) وملاحظة "القفزات" الكبرى في الارتفاع أو ما يعرف بطول الاغصان ، وأيضًا عبر دراسة تغيرات العطالة (Inertie) بين المراحل، والذي يعرف بمقدار الخسارة أو الربح في العطالة اثناء التجميع او العكس. (01 نقاط).

8- **الإجابة:**كلما زادت العطالة الخارجية (بين المجموعات) وانخفضت العطالة الداخلية (داخل المجموعات)، دلّ ذلك على وجود تصنيف فعّال، حيث تكون المجموعات متمايزة ومتجانسة في نفس الوقت. (01 نقاط).

9- **الإجابة:** عندما تكون لدينا فكرة مسبقة عن البنية العامة للمجموعات، أو إذا كنا نرغب في تقسيم مجموعة كبيرة إلى فئات فرعية أكثر دقة، مع المحافظة على التسلسل البنيوي الأصلي. (01 نقاط).

الجانب التطبيقي : 11 نقطة

التمرين الأول: 06 نقاط

الجواب :

1- الخطوات المتبعة هي : (01 نقاط)

الخطوة 1: حساب المسافات الإقليدية (Euclidean Distances) بين مختلف المفردات مثنى مثنى

الخطوة الثانية : تطبيق خوارزمية التصنيف حيث يتم في كل خطوة دمج الزوج الذي يؤدي إلى أقل زيادة ممكنة في **المجموع الكلي للمربعات داخل المجموعات.(Within-cluster variance)** وهذا بدءا من جمع المفردتين الاكثر قربا .

الخطوة الثالثة : ومن أجل حساب المسافات الجديدة يتم اعتماد احد المعايير او المؤشرات كمؤشر وارد أو مؤشر المسافة الدنيا.

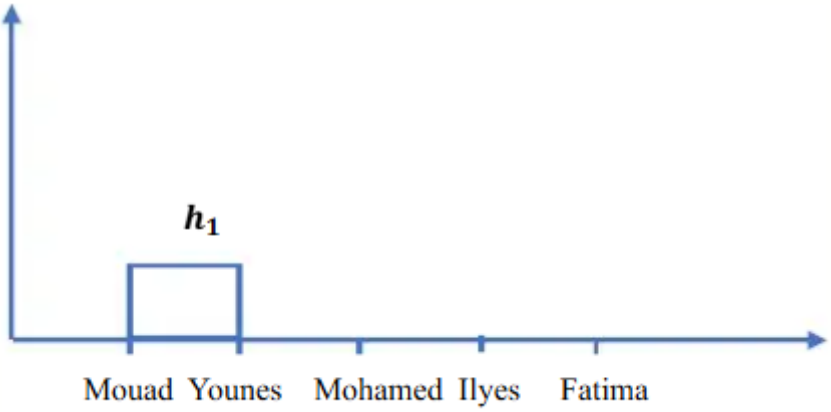
- لا تعطي نتيجة التصنيف نتيجة فريدة من نوعها وانما تختلف باختلاف هذه المؤشرات السابقة الذكر

2- رسم شجرة التصنيف: من أجل ذلك يتم الاعتماد على احد المؤشرات وليكن مؤشر المسافة الدنيا

- **مصفوفة المسافة:** (02 نقاط).

δ^0	Mohamed	Mouad	Younes	Ilyes	Fatima
Mohamed	0				
Mouad	3.16	0			
Younes	3.00	1.00	0		
Ilyes	1.41	4.47	4.12	0	
Fatima	2.83	3.16	2.24	3.16	0

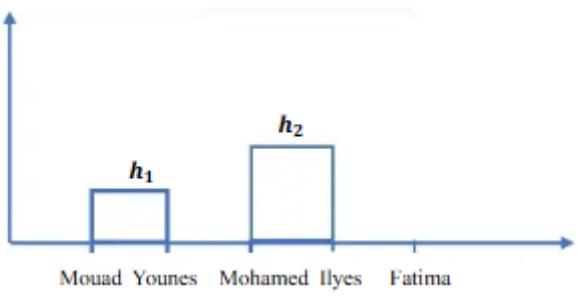
يتم تجميع الفرديتين يونس ومعاذ وذلك باعتبار المسافة بينهما والمقدرة ب 1وهي أقل مسافة مقارنة بالمسافات الاخرى وبعد دمجهما في مجموعة واحدة بحيث يمكن رسم مبدئيا على مستوى شجرة التصنيف و يتم الحصول على مايلي :



ولهذا سوف نقوم بتحيين مصفوفة المسافة بعد اعتمادنا على مؤشر المسافة الدنيا لتصبح على الشكل التالي :

δ^1	Mohamed	Mouad, Younes	Ilyes	Fatima
Mohamed	0			
Mouad, Younes	3.00	0		
Ilyes	1.41	4.12	0	
Fatima	2.83	2.24	3.16	0

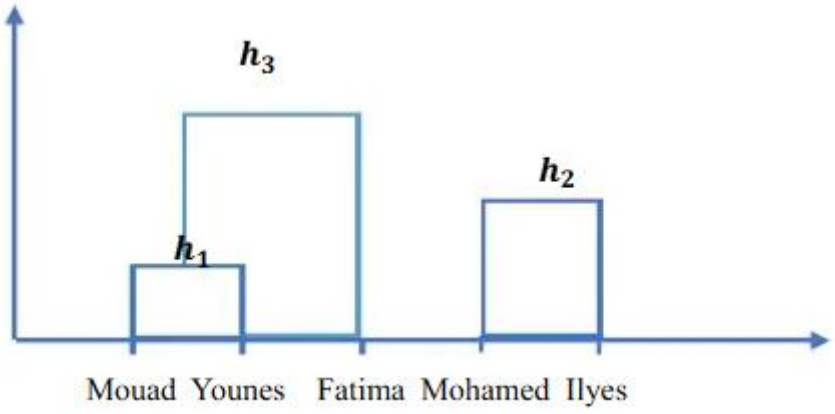
سوف نقوم مرة اخرى من جديد بتجميع المجموعات الاكثر قربا والتي تتمثل في الياس ومحمد . على مستوى شجرة التصنيف يمكننا تمثيل الارتباط الثاني :

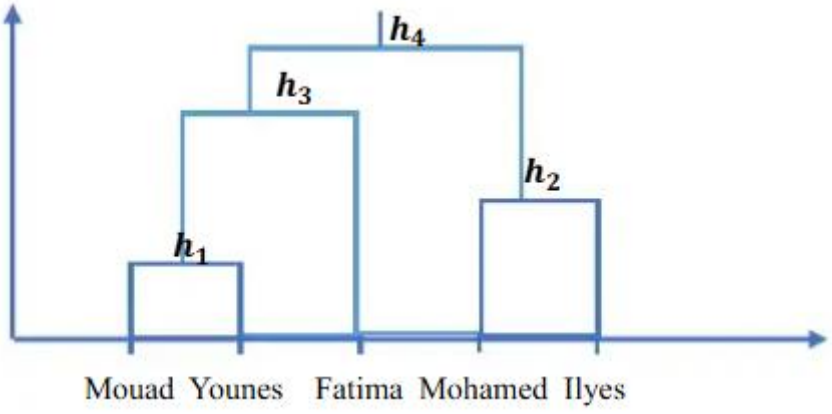


كما يمكننا تحيين مصفوفة المسافة مرة اخرى لتصبح على الشكل :

δ^2	Mouad, Younes	Mohamed, Ilyes	Fatima
Mouad, Younes	0		
Mohamed, Ilyes	3.00	0	
Fatima	2.24	2.83	0

وهكذا يتم الحصول على النتائج الاخرى :





(01 نقاط).

3- (01 نقاط)

الخطوة 1: حساب المركز العام (centre de gravité total)

$$\bar{x} = \left(\frac{5+4+3+2+3}{5}, \frac{4+1+5+5+2}{5}\right) = \left(\frac{17}{5}, \frac{17}{5}\right) = (3.4\ ,3.4)$$

الخطوة 2: حساب العطالة الكلية (Inertie Totale)

نحسب المسافة المربعة لكل فرد إلى المركز العام:

$$I_{totale} = \sum_{i=1}^{n} \frac{1}{n} d^2(x_i, \bar{x})$$

الحسابات:

1. **Mohamed:**

$$(3.4-2)^2 + (3.4-3)^2 = 1.96 + 0.16 = 2.12$$

2. **Mouad:**

$$(3.4-5)^2 + (3.4-2)^2 = 2.56 + 1.96 = 4.52$$

3. **Younes:**

$$(3.4-5)^2 + (3.4-3)^2 = 2.56 + 0.16 = 2.72$$

4. **Ilyes:**

$$(3.4-1)^2 + (3.4-4)^2 = 5.76 + 0.36 = 6.12$$

5. **Fatima:**

$$(3.4-4)^2 + (3.4-5)^2 = 0.36 + 2.56 = 2.92$$

المجموع:

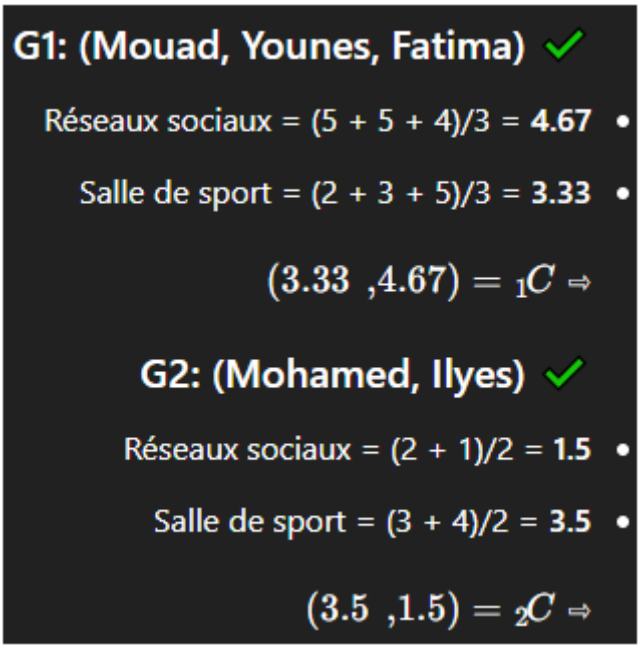
$$I_{totale} = \frac{2.92 + 6.12 + 2.72 + 4.52 + 2.12}{5} = \frac{18.4}{5} = \boxed{3.68}$$

الخطوة 3: افتراض تقسيم الأفراد إلى مجموعتين(Clusters)

لنستخدم نفس التقسيم المقترح مسبقًا: (01 نقاط)

- **G1 شبكات اجتماعية عالية** Mouad, Younes, Fatima
- **G2 شبكات اجتماعية منخفضة** Mohamed, Ilyes

الخطوة 4: حساب مركز كل مجموعة



التمرين الثاني : 05 نقاط

- **1 .المعطيات الأساسية: 0,5 نقاط**

الجدول يتضمن العلاقة بين **الوظيفة** 5فئات:Cadre expérimenté, Cadre débutant, Employé (expérimenté, Employé débutant, Secrétaire) و**درجة التدخين** 4فئات: غير مدخن، مدخن عرضي، متوسط، ومدخن كثيف). وهذا بإجمالي **193 :فردًا**.

2. التحليل العام (القصور) أو العطالة: Inertie – 01 نقاط

التفسير :**البُعد الأول يحمل أغلب المعلومات (حوالي 88 %)**، أي أنه يفسر الفروقات الجوهرية في سلوك التدخين حسب الفئات المهنية.

3. مساهمات الصفوف الفئات المهنية: 01,5 نقاط

مساهمات جزئية في البعدين:

- **Empl_exper**و **Empl_debut** يساهمان بقوة في البعد 1، مما يعني أن هناك اختلافات واضحة في سلوك التدخين بين الموظفين حسب الخبرة.
- **Cadre_debu**و **Cadre_exper**تبرزان في البعد 2، ما يشير إلى تباين داخلي في هذه الفئة.

مساهمات الأعمدة (أنواع المدخنون): 01 نقاط

- (غير المدخنين) و **Moyen**يبرزون في البعد الأول.
- بينما **Occas**و **Gros_fum**أكثر تأثيرًا في البعد الثاني.

5. الاستنتاج العام: 01 نقاط

- **Empl_exper**و **Empl_debut** يرتبطان بشكل أقوى بنمط التدخين، خصوصًا **Empl_exper**الذي يتموضع قرب غير المدخنين، في حين يبدو أن **Cadre_debut**و **Gros_fum**يرتبطان معًا في البعد الثاني.
- **Non_fum**يرتبط بشكل واضح بـ **Empl_exper**.

- **Gros_fum** مرتبط بـ **Cadre_debut** مما يشير إلى أن الموظفين الجدد من الإطارات يميلون إلى التدخين بكثافة أكثر.
- **Secretaire** متوسطة التمثيل، لا تُحدث تمايزًا كبيرًا، لكنها أقرب إلى الفئة."Occas"