

Contrôle intermédiaire ANAD: 2CSSIT et 2CSSIL.
Durée: 2heures

Questions de cours (4pts): Veuillez donner des réponses claires

1- Donner la différence entre l'ACP, L'AFC et l'AFCM. **1pt**

L'ACP: Réduction de dimensionnalité et visualisation des données quantitatives

L'AFC: Analyse de données qualitatives réparties dans un tableau de contingence

L'AFCM: Analyse de questionnaire (ensemble de questions qualitatives posées à n individus):

2- En AFCM que doit t'on éviter pour que les modalités aient des parts d'inerties équivalentes? (voir cours). **0.5pt**

3- En AFCM que signifie la proximité entre deux modalités d'une même variable et celles de deux variables différentes? (Voir cours) **0.5+0.5pt**

4- Supposons qu'on a obtenu ces 6 valeurs propres après une AFCM comment calcule t'on les taux d'inerties modifiés? (Voir cours) **0.5pt**

5- Le tableau de Burt est un tableau:

a- Disjonctif complet

b- Quantitatif

c- **Contingence** **0.25pt**

- Comment est il obtenu? **0.25pt**

$B = Z^t Z$ où Z est un tableau disjonctif.

6- Lorsqu'on effectue l'AFC du tableau de Burt, nous obtenons la projection des individus et des modalités sur les axes factoriels. Vrai ou Faux? Justifiez

Vrai: On obtient ceux des individus en les considérant comme éléments supplémentaires dans le tableau de Burt. **0.5pt**

Exercice1 (10pts): Nous aimerions visualiser à l'aide d'une ACP le tableau de données olympique donnant les performances des athlètes en décathlon (hommes) aux jeux olympiques de 1988. Le décathlon est composé de 10 épreuves: Saut en longueur, Javelot, Perche, Saut en hauteur, 100m, 110m haies, 400m, 1500m, Lancer de disque et Lancer de poids. Quelques résultats de L'ACP donné par R sont donnés ci-après:
Moyennes: 11.19636 7.133333 13.97636 1.982727 49.27667 15.04879 42.35394 4.739394 59.43879 276.0385

Val propres	3.418	2.606	0.943	0.878	0.557	0.491	0.431	0.307	0.267	0.102
% d'inertie	34.182	26.064	9.433	8.780	5.566	4.912	4.306	3.068	2.669	1.019
% d'inertie Cum	34.182	60.246	69.679	78.459	84.026	88.938	93.244	96.312	98.981	100

Variables:

	Dim.1	ctr	cos2	Dim.2	ctr	cos2	Dim.3	ctr	cos2
100	-0.769	17.296	0.591	0.240	2.214	0.058	0.260	7.154	0.067
Long	0.729	15.528	0.531	-0.246	2.313	0.060	0.164	2.854	0.027
Poid	0.498	7.242	0.248	0.781	23.381	0.609	-0.096	0.971	0.009
Haut	0.392	4.506	0.154	0.045	0.078	0.002	0.830	73.100	0.690
400	-0.658	12.663	0.433	0.569	12.402	0.323	0.184	3.591	0.034
110	-0.801	18.791	0.642	0.112	0.484	0.013	0.123	1.592	0.015

Disq	0.325	3.090	0.106	0.813	25.335	0.660	-0.045	0.213	0.002
Perc	0.710	14.752	0.504	0.241	2.237	0.058	-0.133	1.873	0.018
Jave	0.333	3.238	0.111	0.600	13.835	0.361	0.187	3.699	0.035
1500	-0.315	2.895	0.099	0.680	17.721	0.462	-0.216	4.953	0.047

Questions:

1- D'après vous, est ce que les données ont été normalisées? Justifiez. **0.5pt**

Réponse: Oui, car les performances dans les épreuves ont des unités de mesures différentes (100 (sec), Disque(m),....)

2- Quel est le nombre d'axes à retenir? Quelles sont les qualités de représentation du premier et du deuxième plan factoriel. **1pt**

Réponse: On retient 5 axes correspondant à un taux d'inerties cumulées de 84.026%

$$I_{12} = \frac{\lambda_1 + \lambda_2}{\sum \lambda_\alpha} \times 100 = 60,24\%, \quad I_{13} = \frac{\lambda_1 + \lambda_3}{\sum \lambda_\alpha} \times 100 = 43.61\%.$$

3- Quelles sont les variables qui contribuent le plus à la construction de l'axe 1? Donner une signification à cet axe.

Réponse : Les variables qui contribuent le plus sont:

110(-), 100(-), Long(+), Perche (+): **0.5pt**

Cet axe oppose les performances des athlètes en sprint (100 et 110 m haies) à celles des athlètes en saut en longueurs et perche. **1pt**

4- Quelles sont les variables qui contribuent le plus à la construction de l'axe 2? Donner une signification à cet axe.

Réponse : Les variables qui contribuent le plus sont:

Poids (+), Disq(+): **0.5pt**

Cet axe mesure les performances des athlètes en lancer de poids et de disque. **1pt**

5- Quelles sont les variables qui contribuent le plus à la construction de l'axe 3? Donner une signification à cet axe.

Réponse : Les variables qui contribuent le plus sont:

Haut(+): **0.5pt**

Cet axe mesure les performances des athlètes en saut en hauteur (Haut). **0.5pt**

6- Quelles sont les variables qui sont bien représentées par le 1er plan? **1pt**

Nous calculons la contribution relative de chaque variable sur le plan par:

$$C_{re}(j) = \cos^2_1(j) + \cos^2_2(j).$$

Le résultat est:

100	Long	poids	haut	400	110	disq	Perc	jave	1500
0.648	0.591	0.856	0.156	0.756	0.654	0.765	0.562	0.471	0.560

Donc les variables les mieux représentées par le plan sont: Poids, 400, disq. **0.5pt**

7- Par quel plan , la variable "Haut" est- elle bien représentée?

Sur le premier plan sa contribution est de 0.156 (voir la question précédente) **0.25pt**

Sur le deuxième plan (1,3): $C_{re}(haut) = \cos_1^2(haut) + \cos_3^2(haut) = 0.15403795 + 0.689551577 = 0.8435895$, **0.25pt**

Sur le troisième plan (2,3): $C_{re}(haut) = \cos_2^2(haut) + \cos_3^2(haut) = 0.689551577 + 0.002 = 0.691$ **0.25pt.**

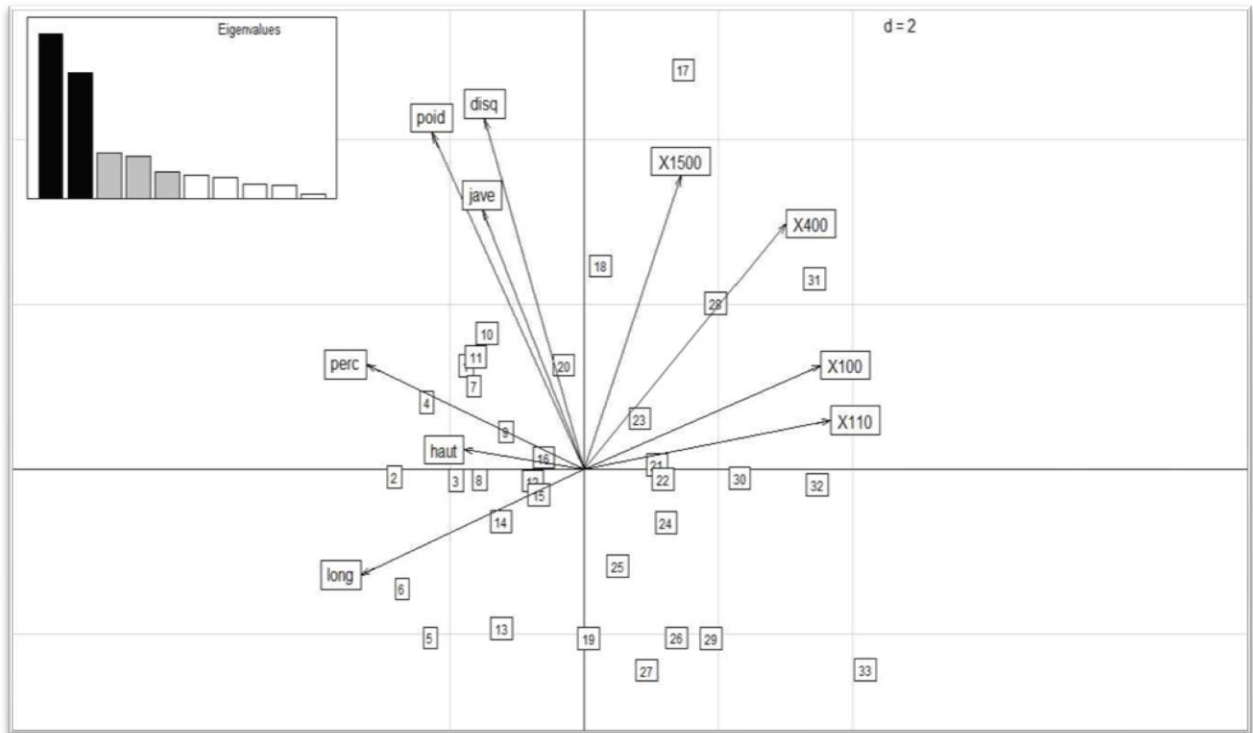
Elle est largement donc bien représentée sur le deuxième plan factoriel. **0.25pt**

8- Selon la représentation des individus/variables, qui sont les athlètes qui contribuent le plus à la construction de l'axe1, caractérisez les. Même question pour l'axe 2.

Réponse:

Axe 1: Ceux qui contribuent le plus sont ceux qui sont les plus éloignés de l'origine: Ind2, Ind6, Ind 4 qui ont de bonnes performances en saut (longueur et perche). et Ind 32, Ind 31, Ind33 qui sont plutôt bon en sprint . **1pt**

Axe 2: Ceux qui contribuent le plus sont ceux qui sont les plus éloignés de l'origine: Ind17, Ind18, qui ont de bonnes performances en lancer de poids et disque et Ind 19, Ind 27 qui ont des performances au dessous de la moyenne dans ces deux épreuves. **1pt**



Exercice2 (6pts) : Soit la répartition de 47 individus selon leur réponses à 2 questions; X=" Le programme réalisé correspond au programme annoncé" et Y=" Le volume horaire prévu dans l'UE a été respecté". Les réponses (modalités des deux variables sont: Tout à fait d'accord (TD), D'accord (D), Pas d'accord (PaD), En total désaccord (TDS)).

	Q1TD	Q1D	Q1PaD	Q1TDS
Q2TD	10	5	1	0
Q2D	5	13	2	0
Q2PaD	0	2	4	1
Q2TDS	0	0	0	4

Les résultats de l'AFC avec le logiciel R sont comme suit:

Vps 0.822 **xxx** 0.108
 % d'inertie. 67.305 23.865 **xxx**
 % d'inertie Cumulée 67.305 91.171 100.000

Résultats ligne:

	Dim.1	ctr	cos2	Dim.2	ctr	cos2	Dim.3	ctr	cos2
Q2_TD	-0.411	6.995	0.352	-0.443	xxx	0.410	0.338	36.011	0.238
Q2_D	-0.355	6.517	0.492	0.036	0.186	0.005	-0.359	50.744	0.238
Q2_PaD	0.319	1.847	0.065	1.176	70.618	0.877	0.303	12.642	0.058
Q2_TDS	2.860	84.642	0.974	-0.463	6.245	0.025	-0.087	0.603	0.001

Résultats Colonne

	Dim.1	ctr	cos2	Dim.2	ctr	cos2	Dim.3	ctr	cos2
Q1TD	-0.433	7.264	0.330	-0.525	30.211	0.487	0.322	30.610	0.183
Q1D	-0.332	xxx	0.454	0.055	0.448	0.013	-0.361	51.278	0.534
Q1PaD	0.025	0.011	0.000	1.146	67.065	0.909	0.361	18.031	0.090
Q1TDS	2.593	87.004	0.991	-0.250	2.276	0.009	-0.029	0.081	0.000

Questions:

1- Calculer la distance du KHI2 entre les 2 profils colonnes Q1TD et Q1D. Calculer la distance entre leur projections. Que constatez vous?

$$d_{KHI2}^2(Q1TD, Q1D) = \sum_i \frac{1}{f_i} \left(\frac{K_{i1}}{K_1} - \frac{K_{i2}}{K_2} \right)^2 = \frac{47}{16} \left(\frac{10}{15} - \frac{5}{20} \right)^2 + \dots + \frac{47}{4} \left(\frac{0}{15} - \frac{0}{20} \right)^2 = 0.509 + 0.235 + 0.0671 = 0.811 \quad \text{0.75pt}$$

$$d^2(V(Q1TD), V(Q1D)) = (-0.433 + 0.332)^2 + (-0.525 - 0.055)^2 = 0.346 \quad \text{0.5pt}$$

Les deux distances sont différentes ce qui indique qu'il y a une perte d'information concernant la proximité de ces deux modalités dans la représentation faite sur le 1er plan factoriel (1,2). **0.25pt**

2- Donner les poids des PLS et des PCS.

Lignes **0.25pt**

Q2_TD	Q2_D	Q2_PaD	Q2_TDS
0.34042553	0.42553191	0.14893617	0.08510638

Colonnes **0.25pt**

Q1TD	Q1D	Q1PaD	Q1TDS
0.3191489	0.4255319	0.1489362	0.1063830

3- Compléter les vides. **1.25pt**

Nous avons $I_2 = \frac{\lambda_2}{\sum \lambda_\alpha} \Rightarrow \lambda_2 = \frac{\lambda_1 + \lambda_3}{1 - I_2} = 0.292$, $I_2 = \frac{\lambda_3}{\sum \lambda_\alpha} \times 100 = I_{123} - I_{12} = 13,829\%$,

$$C_{ab}^2(Q2TD) = f_{1.} \frac{Y_2^2(Q2TD)}{\lambda_2} = 0.186, C_{ab}^1(Q1D) = f_{1.} \frac{Y_1^2(Q1D)}{\lambda_1} = 5,720.$$

$$C_{re}^3(Q1TDS) = \frac{V_3^2(Q1TDS)}{d_{KHI2}^2(Q1TDS, g_p)} = \frac{(0.029)^2}{d_{KHI2}^2(Q1TDS, g_p)} = 0.0001238$$

telle que $d_{KHI2}^2(Q1TDS, g_p) = \sum_i \frac{1}{f_i} \left(\frac{K_{i4}}{K_4} - f_i \right)^2$

4- Donner une signification à l'axe1 et à l'axe2.

Nous comparons les contributions absolues des profils à leurs poids, pour ressortir ceux qui contribuent le plus:

Axe1: **1pt**

PLs	Cs
Q1TDS (+)	Q2TDS(+)

signification: l'axe 1 représente ceux qui sont ont total désaccord sur le fait que le programme réalisé correspond à l'énoncé et que le volume horaire est respecté.

Axe2: **1pt**

PLs	Cs
Q1PaD (+)	Q2PaDS(+)

signification: l'axe 2 représente ceux qui ne sont pas d'accord sur le fait que le programme réalisé correspond à l'énoncé et que le volume horaire est respecté.

5- Soit une ligne supplémentaire (1,4,0,3) , quelle est sa projection sur l'axe1. **0.5pt**

$$Y_1(5) = \frac{1}{\sqrt{\lambda_1}} \sum \frac{K_{5j}}{K_5} V_1(j) = \frac{1}{\sqrt{0.822}} \sum \frac{K_{5j}}{8} V_1(j) = \frac{0.75}{\sqrt{0.822}} = 0.827$$

6- Quelles sont les lignes et les colonnes les mieux représentées et par quels axes. **0.5pt**

Par l'axe 1: Q2TDS et Q1TDS, par l'axe2: Q2PaD et Q1PaD, car leurs contributions relatives sur les axes respectives sont proches de 1.

Bon courage