

Les étudiants sont autorisés à utiliser une machine à calculer et à consulter les notes de cours, sous forme de transparents distribués par l'enseignant, sans ajout de notes personnelles

I Questions de cours (4 points):

1. Expliquer brièvement comment on peut construire un indicateur synthétique d'une dimension particulière au moyen de l'Analyse des Correspondances Multiples (2 points).
2. Expliquer brièvement la différence existante entre classification et classement des individus (1 point).
3. Définir le point moyen d'un tableau de données noté X, représentation de l'ensemble des valeurs prises par p variables continues pour un échantillon de n individus (1 point).

Exercice 1 (7,5 points):

L'objectif est d'analyser les données techniques de 24 modèles de voiture vendus en France en 2004. Les variables considérées sont : Cylindrée (cylindrée, en cm³), Puissance (puissance, en chevaux fiscaux), Vitesse (vitesse maximum en km/h), Poids (poids total en kg), Largeur (largeur de la voiture en millimètres), Longueur (longueur de la voiture en millimètres) Les individus étudiés sont les 24 modèles des marques Mercedes, Nissan, Audi, Jaguar, Alfa-Romeo, BMW, Aston-Martin, Bentley, Ferrari, Volkswagen, Land Rover, Austin, Citroën, Fiat, Peugeot et Renault. La matrice des coefficients de corrélation linéaire entre les variables est la suivante :

	Cylindrée	Puissance	Vitesse	Poids	Largeur	Longueur
Cylindrée	1.000	0.954	0.885	0.692	0.706	0.664
Puissance	0.954	1.000	0.934	0.529	0.730	0.527
Vitesse	0.885	0.934	1.000	0.466	0.619	0.578
Poids	0.692	0.529	0.466	1.000	0.477	0.795
Largeur	0.706	0.730	0.619	0.477	1.000	0.591
Longueur	0.664	0.527	0.578	0.795	0.591	1.000

1. Pourquoi est-il nécessaire ici de remplacer les données d'origine par les données centrées-réduites (1 point)

On réalise une Analyse en Composantes Principales sur les variables centrées et réduites. On donne ci-après les valeurs propres associées aux composantes principales, sauf la valeur de la troisième (en gras ci-dessous).

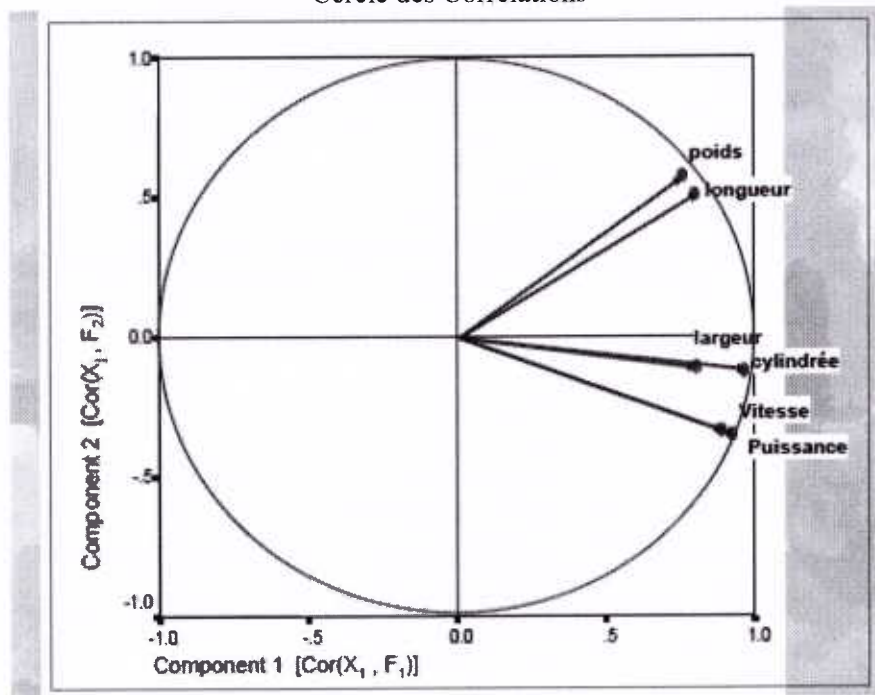
Valeurs propres :

Axes	1	2	3	4	5	6
Valeurs Propres	4,411	0,853	Xxx	0,236	0,051	0,012

2. Définir et déterminer l'inertie totale du nuage des données traité. Expliquer comment on peut en déduire la valeur manquante de la valeur propre associée au troisième axe ? (2 points)
3. Combien d'axes doit-on garder pour l'analyse ? Quelle sera alors la qualité globale de la représentation du nuage sur ce sous ensemble d'axes ? (1,5 point)
4. Le tableau ci-dessous donne les coefficients de corrélation entre les six variables centrées réduites et les deux premières composantes principales. A l'aide de ce tableau, et du graphique du cercle des corrélations, expliquer le type d'effet que nous observons sur le premier axe (1 point).

Corrélations variables / composantes principales		
	1 ^{ère} comp	2 ^{ème} comp
Cylindrée	0.97	-0.17
Puissance	0.95	-0.42
Vitesse	0.94	-0.40
Poids	0.74	0.62
Largeur	0.83	-0.15
Longueur	0.81	0.60

Cercle des Corrélations



5. Expliquer les variables qui déterminent les deux premières composantes principales (précisez clairement les critères utilisés). Interpréter ces deux composantes (2 points).

Exercice 2 (8,5 points):

L'objectif est d'analyser les avis d'un ensemble de personnes interrogées au sujet des OGM (Organismes Génétiquement Modifiés). 135 personnes ont été contactées pour un entretien en face à face. Les variables issues des questions et les modalités associées sont les suivantes :

Concerné par les OGM : Pas du tout, Un peu, Moyennement, Beaucoup.

Position sur culture des OGM : Pas favorable, Plutôt défavorable, Favorable, Très favorable.

Position sur alimentation avec les OGM : Pas favorable, Plutôt défavorable, Favorable, Très favorable

Sont-ils utiles pour prévenir la famine : Non, Oui.

Constituent-ils une amélioration pour l'agriculture : Non, Oui.

Constituent-ils un progrès dans le futur : Non, Oui.

Constituent-ils un danger pour l'homme : Non, Oui.

Constituent-ils une menace pour la planète : Non, Oui.

Est-ce un procédé inutile : Non, Oui.

Sexe : Femme, Homme.

CSP : Cadre, Fonctionnaire, Technicien, Ouvrier, Commerçant, Profession libérale, Retraité, Inactif, Etudiant, Autre.

Orientation Politique : Droite, Centre, Gauche, Verts, Extrême Gauche.

1. On réalise une Analyse des Correspondances Multiples (ACM) sur les 135 personnes, en utilisant comme variables actives les 9 premières variables. Compte tenu des modalités listées ci-dessus,
 - a. Déterminer l'inertie totale du nuage caractérisé par ces 9 variables (0,5 point).
 - b. Connaissant les modalités de chacune des 9 variables, calculer la contribution des deux premières variables actives (1 point).
 - c. De ce point de vue, est-ce que les modalités choisies pour ces neuf variables sont susceptibles de présenter un problème pour l'ACM (0,5 point) ?
2. Les effectifs des 3 premières variables actives sont donnés dans les tableaux suivants :

Concerné	Fréquence
Beaucoup	36
Moyen	53
Un Peu	31
Pas du tout	15

Position_Culture	Fréquence
Très Favorable	3
Favorable	45
Plutôt Défavorable	54
Pas Favorable du Tout	33

Position Alimentation	Fréquence
Très Favorable	1
Favorable	37
Plutôt Défavorable	47
Pas Favorable du Tout	50

- a. Expliquer ce que l'on devrait modifier dans la constitution de certaines modalités de ces trois variables de manière à éviter des problèmes d'interprétation de l'ACM (1 point).
 - b. En déduire la nouvelle inertie du nouveau tableau de données tel que vous proposez de le modifier (0,5 point)
3. Après modification de la constitution de deux variables actives, qui ont perdu chacune une modalité, une ACM est réalisée dont on présente un extrait de l'éboulis des valeurs propres:

N	Val.Pr.	Diff.	Pct	Cum	!
1	0.3219	.	22.99	22.99	
!*****					

2	0.1931	0.1288	13.80	36.79	
!*****					
3	0.1278	0.0653	9.13	45.92	!*****
4	0.1058	0.0221	7.56	53.48	!*****
5	0.0980	0.0078	7.00	60.47	!*****
6	0.0931	0.0049	6.65	67.12	!*****
7	0.0845	0.0085	6.04	73.16	!*****

a. Combien de valeurs propres non trivialement égales à 0 ou 1 sont engendrées par cette ACM ?
Combien d'axes principaux pouvons-nous choisir d'interpréter ? (1 point)

4. Les aides à l'interprétation sur les deux premiers axes sont données ci-dessous :

Modalité	POIDS	Axe 1				Axe 2			
		COORD	CTR	CO2	QLT	COORD	CTR	CO2	QLT
Concerné beaucoup	2.67	-0.84	5.9	25.8	25.8	0.54	4.0	10.4	36.2
Concerné moyennement	3.93	0.21	0.5	2.9	2.9	-0.03	0.0	0.1	3.0
Concerné peu	2.30	0.24	0.4	1.7	1.7	-0.85	8.6	21.5	23.2
Concerné pas du tout	1.11	0.79	2.1	7.8	7.8	0.59	2.0	4.4	12.2
Culture favorable	3.56	0.98	10.5	52.7	52.7	0.56	5.8	17.5	70.2
Culture plutôt défavorab.	4.00	-0.06	0.0	0.2	0.2	-0.93	17.9	57.7	57.9
Culture pas favor. du tout	2.44	-1.32	13.3	56.6	56.6	0.70	6.3	16.0	72.5
Alimentation favorable	2.81	1.10	10.5	47.1	47.1	0.87	11.0	29.5	76.6
Alimentation plutôt défav.	3.48	0.16	0.3	1.4	1.4	-1.05	19.8	58.6	60.0
Alimentation pas fav du tout	3.70	-0.98	11.1	56.9	56.9	0.33	2.0	6.2	63.1
Prévention Famine non	4.96	-0.56	4.9	31.3	31.3	0.01	0.0	0.0	31.3
Prévention Famine oui	5.04	0.55	4.8	31.3	31.3	-0.01	0.0	0.0	31.3
Améliore. Agriculture non	6.89	-0.20	0.8	8.5	8.5	0.11	0.4	2.6	11.1
Améliore. Agriculture oui	3.11	0.43	1.8	8.5	8.5	-0.24	0.9	2.6	11.1
Progrès non	4.00	-0.43	2.2	12.1	12.1	-0.09	0.2	0.5	12.6
Progrès oui	6.00	0.28	1.5	12.1	12.1	0.06	0.1	0.5	12.6
Danger homme non	2.89	0.89	7.1	32.3	32.3	0.68	6.9	18.8	51.0
Danger homme oui	7.11	-0.36	2.9	32.3	32.3	-0.28	2.8	18.8	51.0
Menace planète non	3.56	0.77	6.6	32.9	32.9	0.06	0.1	0.2	33.1
Menace planète oui	6.44	-0.43	3.6	32.9	32.9	-0.03	0.0	0.2	33.1
Procédé inutile non	9.11	0.16	0.7	24.7	24.7	-0.08	0.3	7.0	31.7
Procédé inutile oui	0.89	-1.59	7.0	24.7	24.7	0.85	3.3	7.0	31.7

Les abréviations signifient : COORD : coordonnées, CTR : contribution, CO2 : cosinus carré de l'angle avec l'axe, QLT : qualité de représentation

- a. D'après ce tableau, donnez les modalités qui contribuent significativement aux deux axes (vous expliquerez comment vous faites votre raisonnement) (1 point)
 - b. Interprétez ces deux premiers axes (1 point)
5. Les coordonnées des variables supplémentaires sexe et parti politique dans ce premier plan factoriel, ainsi que leurs valeurs test d'interprétation, sont données dans ce tableau ci-dessous:

Modalité	Axe 1					Axe 2				
	POIDS	COORD	CO2	QLT	Valeur Test	COORD	CO2	QLT	Valeur Test	
Centre non	76.30	-0.08	2.0	2.0	-1.64	-0.01	0.0	2.0	-0.11	
Centre oui	23.70	0.25	2.0	2.0	1.64	0.02	0.0	2.0	0.11	
Ext. Gauche non	93.33	0.07	6.6	6.6	2.97	0.01	0.1	6.7	0.26	
Ext. Gauche oui	6.67	-0.96	6.6	6.6	-2.97	-0.09	0.1	6.7	-0.26	
Gauche non	65.19	0.16	4.9	4.9	2.57	0.05	0.4	5.3	0.75	
Gauche oui	34.81	-0.30	4.9	4.9	-2.57	-0.09	0.4	5.3	-0.75	
Droite non	70.37	-0.24	13.5	13.5	-4.25	-0.01	0.0	13.5	-0.18	
Droite oui	29.63	0.57	13.5	13.5	4.25	0.02	0.0	13.5	0.18	
Verts non	94.81	0.06	6.9	6.9	3.05	-0.03	1.3	8.2	-1.34	
Verts oui	5.19	-1.13	6.9	6.9	-3.05	0.50	1.3	8.2	1.34	
Femme non	47.41	-0.01	0.0	0.1	-0.11	0.21	3.9	4.0	2.30	
Femme oui	52.59	0.01	0.0	0.1	0.11	-0.19	3.9	4.0	-2.30	

- a. Interpréter la signification des deux premiers axes au regard de la position des variables supplémentaires (1 point).
6. Expliquez comment vous pourriez à l'aide des résultats de cette ACM réaliser une typologie des individus en fonction de leurs opinions au sujet des OGM. (1 point)