

**EXAMEN DE RATTRAPAGE D'ANALYSE DES DONNEES P**

**NB :** Aucun document n'est autorisé

Durée : 1h @0mm

**Exercice 1 : Question de cours**

**Question 1 :** L'Analyse en Composantes Principales (ACP) vise essentiellement quatre (04) objectifs en liaison avec les individus et/ou les variables puis le tableau brut des données. Rappelez ces quatre (04) objectifs de l'ACP.

**Question 2 :** En pratique, dans une ACP, on préfère travailler sur des variables centrées et réduites. Quel intérêt présente ce procédé et quelles sont les propriétés qui s'en découlent relativement au nuage de points des individus ?

**Question 3 :** On considère un tableau de valeur numérique X ayant n lignes et p colonnes. Dans la définition des composantes principales, on est amené à chercher des sous-espaces vectoriel à une dimension (droite passant par l'origine) et ensuite à deux dimension, qui réalise le meilleur ajustement possible du nuage de points.

En désignant par  $u_1$  le vecteur directeur unitaire de la première droite d'ajustement et toujours par  $u_1$  la matrice colonne à lui associée, montrer à travers la résolution du programme :

$$\begin{cases} \max u_1' X' X u_1 & s/c \\ \|u_1\|^2 = 1 \end{cases}$$

Que  $u_1$  est le vecteur propre associé à la plus grande valeur propre de la matrice  $X'X$

**Question 4 :** Dans le but de résumer l'information contenue dans les données, l'ACP ne s'intéresse qu'aux points saillants qui structurent les données. De ce fait, on se limite à l'interprétation de quelques axes. Plusieurs règles empiriques aident à déterminer le nombre d'axes significatifs à retenir. Enumérez les deux(02) plus importantes que vous connaissez et énoncer-les.

**Question 5 :** Les relations de transition en ACP montrent qu'à tout vecteur propre  $u_\alpha$  de la matrice  $X'X$  relatif à une valeur propre  $\lambda_\alpha$  non nulle, correspond un vecteur propre  $Xu_\alpha$  de  $XX'$  relatif à la même valeur propre  $\lambda_\alpha$ .

En désignant par  $v_\alpha$  le vecteur propre de  $XX'$  correspondant à la valeur propre  $\lambda_\alpha$ , donner les formules de transition entre les deux espaces  $R^p$  des individus et l'espace  $R^n$  des variables.

**Exercice 2 : Une Analyse Factorielle des Correspondances (AFC)**

i) On considère le tableau suivant croisant le sexe et le niveau d'instruction

	Niveau d'instruction				Total
	Aucun	Primaire	Secondaire	Supérieur	
Masculin	25	30	40	35	130
Féminin	45	28	12	10	95
Total	70	58	52	45	225

- 1) a. Construire le tableau des profils-lignes.
- b. Construire le tableau des profils-colonnes.
- c. Construire le tableau des fréquences observées.
- d. Construire le tableau des fréquences théoriques.
- 2) a. Sous quelle condition les deux variables seront-elles indépendantes ?
- b. Donner la formule de la statistique de Khi-deux (Khi-carré) et calculer sa valeur à partir des données de l'exercice.

3) a. En comparant cette valeur calculée à celle d'un Khi-deux tabulé au seuil de 5% (voir table en annexe 2) dont vous préciserez le nombre de degrés de liberté, se prononcer sur l'indépendance des deux variables.

b. L'AFC se justifie-t-elle ? Pourquoi ?

ii)

- 1) Qu'est-ce qui justifie l'utilisation de la métrique de khi-deux pour mesurer la distance entre deux profils en Analyse Factorielle des Correspondances (AFC) ?
- 2) Montrer que le centre de gravité du nuage des profil-ligne est  $g_l = f_j$
- 3) Montrer qu'en AFC, l'inertie des profils-lignes par rapport à leur centre de gravité ( $I_N$ ) est égale à  $\frac{\chi^2}{n}$ , où  $\chi^2$  désigne le khi-deux.

Indication :  $I_N = \sum_l f_l d^2(l, g_l)$  avec g le centre de gravité du nuage

4) La contribution d'un individu i à la formation d'un axe  $\alpha$  est calculée en ACP par :

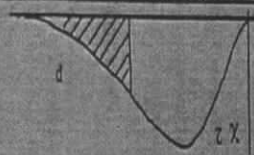
$$CTR_{\alpha}(i) = \frac{F_{\alpha}^2(i)}{\lambda_{\alpha}} \text{ où } F_{\alpha} \text{ désigne la coordonnée de l'individu et } \lambda_{\alpha} = \sum_l F_{\alpha}^2(l).$$

Qu'en est-il en AFC ? Rappelez le critère d'identification des individus qui contribuent le plus à la formation d'un axe  $\alpha$  en ACP ?

ANNEXE : Table de Khi-deux

TABLE DU CHI-DEUX :  $\chi^2(n)$

0.01	0.02	0.05	0.10	0.20	0.30	0.50	0.70	0.80	0.90	$n$
6.635	7.879	9.841	11.668	13.277	15.086	16.812	18.475	20.090	21.666	1
9.210	10.597	12.592	14.664	16.266	18.001	19.591	21.026	22.454	23.664	2
11.341	12.838	15.086	17.179	18.907	20.515	22.161	23.589	24.786	26.179	3
13.277	14.664	17.151	19.023	21.026	22.161	23.589	24.786	26.179	27.657	4
15.086	16.266	18.467	20.290	22.307	23.337	24.433	25.564	26.751	28.341	5
16.812	17.779	19.758	21.454	23.337	24.433	25.564	26.751	27.981	29.288	6
18.475	19.023	20.891	22.454	24.290	25.188	26.286	27.204	28.187	30.191	7
20.090	20.290	21.990	23.337	25.188	26.286	27.204	28.187	29.154	31.024	8
21.666	21.454	23.026	24.290	26.179	27.204	28.187	29.154	30.191	31.902	9
23.209	22.618	24.026	25.188	27.204	28.187	29.154	30.191	31.171	32.909	10
24.725	23.664	25.000	26.179	28.187	29.154	30.191	31.171	32.171	33.902	11
26.217	24.654	25.927	27.204	29.154	30.191	31.171	32.171	33.171	34.805	12
27.688	25.612	26.801	28.187	30.191	31.171	32.171	33.171	34.171	35.705	13
29.141	26.541	27.622	29.154	31.171	32.171	33.171	34.171	35.171	36.592	14
30.578	27.442	28.393	30.191	32.171	33.171	34.171	35.171	36.171	37.466	15
32.000	28.312	29.133	31.171	33.171	34.171	35.171	36.171	37.171	38.328	16
33.409	29.152	29.843	32.171	34.171	35.171	36.171	37.171	38.171	39.179	17
34.805	29.964	30.521	33.171	35.171	36.171	37.171	38.171	39.171	40.017	18
36.191	30.747	31.171	34.171	36.171	37.171	38.171	39.171	40.171	40.841	19
37.566	31.501	31.843	35.171	37.171	38.171	39.171	40.171	41.171	41.651	20
38.932	32.226	32.483	36.171	38.171	39.171	40.171	41.171	42.171	42.446	21
40.289	32.923	33.094	37.171	39.171	40.171	41.171	42.171	43.171	43.226	22
41.638	33.592	33.674	38.171	40.171	41.171	42.171	43.171	44.171	44.000	23
42.980	34.232	34.224	39.171	41.171	42.171	43.171	44.171	45.171	44.768	24
44.314	34.843	34.745	40.171	42.171	43.171	44.171	45.171	46.171	45.529	25
45.642	35.425	35.236	41.171	43.171	44.171	45.171	46.171	47.171	46.282	26
46.963	35.978	35.697	42.171	44.171	45.171	46.171	47.171	48.171	47.027	27
48.278	36.503	36.128	43.171	45.171	46.171	47.171	48.171	49.171	47.764	28
49.588	37.000	36.529	44.171	46.171	47.171	48.171	49.171	50.171	48.492	29
50.892	37.469	36.899	45.171	47.171	48.171	49.171	50.171	51.171	49.212	30



Pour  $n > 30$ , on peut admettre que  $\sqrt{2\chi^2} - \sqrt{2n-1} \sim N(0,1)$