

Université PANTHEON – ASSAS (PARIS II)

Droit-Economie-Sciences Sociales

Septembre 2019

Master 1 Economie-Gestion mention Ingénierie économique et statistique

Analyse des données

(Unité d'Enseignements Fondamentaux 1)

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 (4.5 points)

1. Expliquer les différences existant entre les méthodes de classification hiérarchique et non hiérarchique (1.5 point)
2. Définir le point moyen d'un tableau de données noté X , représentation de l'ensemble des valeurs prises pour p variables par n individus (1 point).
3. Expliquer comment on peut construire un indicateur synthétique d'une dimension particulière présente dans les données au moyen de l'Analyse des Correspondances Multiples. Vous pourrez vous appuyer sur un exemple (2 points).

Exercice 1 (10,5 points):

Le tableau ci-dessous représente les performances des grands pays européens dans les disciplines de l'athlétisme. Il comptabilise les titres de champion d'Europe selon la nationalité des athlètes masculins et féminins vainqueurs.. Les pays sont respectivement l'Allemagne, la Belgique, l'Espagne, la France, la Grande-Bretagne (GB), l'Italie, les Pays Bas et la Russie. Les disciplines récompensées sont la chimie, l'économie, la littérature, la médecine, la paix et la physique.

Le tableau suivant indique, pour chaque pays, le nombre de fois où un ou une athlète a été le champion d'Europe dans une discipline particulière, sachant que plusieurs athlètes peuvent partager les titres.

	Sprint	Demi Fond	Fond	Lancers	Saut	Relais
Allemagne	24	1	8	18	5	24
Belgique	4	3	5	2	3	10
Espagne	4	3	2	4	1	4
France	8	3	11	12	10	9
GB	23	6	7	26	11	20
Italie	1	1	6	5	1	5
Pays Bas	6	0	2	3	1	11
Russie	51	43	8	70	19	66

1. Répondez aux deux questions suivantes (1 point):

- a. Combien de titres de champions d'Europe ont été obtenus dans ces huit pays?
- b. Combien d'individus belges ont reçu un titre de champion d'Europe ?

2. A l'aide des tableaux des profils lignes et des profils colonnes ci-dessous, déterminez la probabilité que le titre de champion d'Europe soit en lancers si le vainqueur est un français. Ensuite, déterminez la probabilité que le vainqueur soit français lorsqu'il a obtenu son titre en sprint. (1 point).

Profils de lignes

	Sprint	Demi-Fond	Fond	Lancers	Sauts	Relais
Allemagne	0.300000	0.012500	0.100000	0.225000	0.062500	0.300000
Belgique	0.148148	0.111111	0.185185	0.074074	0.111111	0.370370
Espagne	0.222222	0.166667	0.111111	0.222222	0.055556	0.222222
France	0.150943	0.056604	0.207547	0.226415	0.188679	0.169811
GB	0.247312	0.064516	0.075269	0.279570	0.118280	0.215054
Italie	0.052632	0.052632	0.315789	0.263158	0.052632	0.263158
Pays Bas	0.260870	0.000000	0.086957	0.130435	0.043478	0.478261
Russie	0.198444	0.167315	0.031128	0.272374	0.073930	0.256809

Profils de colonnes

	Sprint	Demi-Fond	Fond	Lancers	Sauts	Relais
Allemagne	0.198347	0.016667	0.163265	0.128571	0.098039	0.161074
Belgique	0.033058	0.050000	0.102041	0.014286	0.058824	0.067114
Espagne	0.033058	0.050000	0.040816	0.028571	0.019608	0.026846
France	0.066116	0.050000	0.224490	0.085714	0.196078	0.060403
GB	0.190083	0.100000	0.142857	0.185714	0.215686	0.134228
Italie	0.008264	0.016667	0.122449	0.035714	0.019608	0.033557
Pays Bas	0.049587	0.000000	0.040816	0.021429	0.019608	0.073826
Russie	0.421488	0.716667	0.163265	0.500000	0.372549	0.442953

3. Compte tenu de la forme des données recueillies, quelle type d'analyse factorielle pouvons-nous réaliser (0,5 point)? Après celle-ci, le tableau de résultats donnant l'inertie et les valeurs propres pour chaque axe est obtenu ci-dessous. Pourquoi ce tableau ne contient-il que cinq valeurs propres ? D'après ces résultats, combien d'axes devrions-nous retenir pour l'interprétation ? (2 points)

Tableau de l'inertie des axes

Valeur Propre	Inertie principale	Khi-2	Pourcentage	Pourcent. cumulé
0.28867	0.08333	47.4988	54.75	54.75
0.19350	0.03744	21.3425	24.60	79.35
0.14716	0.02166	12.3434	14.23	93.57
0.08879	0.00788	4.4940	5.18	98.75
0.04354	0.00190	1.0805	1.25	100.00
Total	0.15221	86.7592	100.00	

4. Quels sont les pays qui caractérisent les deux premiers axes principaux (vous justifierez précisément votre réponse à l'aide du tableau d'aide à l'interprétation pour les modalités ligne actives) (1.5 points).

Aides à l'interprétation pour les modalités-lignes actives

!	!				AXE1				!				AXE2				!
!	CONTR	POIDS	!	COORD	CTR	RCTR	CO2	QLT	!	COORD	CTR	RCTR	CO2	QLT	!		
!	Ident.		!						!						!		
!	Allemagne	12.53	14.04	!	0.16	4.2	5	18.5	18.5	!	-0.32	38.3	1	75.2	93.7	!	
!	Belgique	9.48	4.74	!	0.36	7.2	4	41.6	41.6	!	0.05	0.3	8	0.8	42.4	!	
!	Espagne	1.34	3.16	!	-0.05	0.1	8	3.5	3.5	!	0.11	1.1	6	19.9	23.4	!	
!	France	21.73	9.30	!	0.50	27.7	2	69.8	69.8	!	0.27	18.5	3	21.0	90.8	!	
!	GB	4.83	16.32	!	0.04	0.2	7	2.8	2.8	!	-0.04	0.7	7	3.7	6.5	!	
!	Italie	17.02	3.33	!	0.71	20.3	3	65.3	65.3	!	0.27	6.3	4	9.0	74.4	!	
!	Pays Bas	9.91	4.04	!	0.18	1.5	6	8.2	8.2	!	-0.52	28.7	2	71.1	79.3	!	
!	Russie	23.15	45.09	!	-0.27	38.7	1	91.5	91.5	!	0.07	6.1	5	6.5	98.1	!	

5. Quelles sont les disciplines qui caractérisent les deux premiers axes principaux (vous justifierez précisément votre réponse à l'aide du tableau d'aide à l'interprétation pour les modalités ligne actives) (1.5 points).

Aides à l'interprétation pour les modalités-colonnes actives

!	!				AXE1				!				AXE2				!
!	CONTR	POIDS	!	COORD	CTR	RCTR	CO2	QLT	!	COORD	CTR	RCTR	CO2	QLT	!		
!	Sprint	8.85	21.23	!	-0.06	0.9	5	5.3	5.3	!	-0.21	25.5	2	70.8	76.1	!	
!	Demi-Fond	25.56	10.53	!	-0.46	26.9	2	57.7	57.7	!	0.35	34.7	1	33.4	91.0	!	
!	Fond	38.39	8.60	!	0.79	64.4	1	91.8	91.8	!	0.19	7.9	5	5.1	96.9	!	
!	Lancers	6.85	24.56	!	-0.11	3.4	4	27.3	27.3	!	0.07	2.9	6	10.3	37.6	!	
!	Sauts	10.65	8.95	!	0.20	4.4	3	22.8	22.8	!	0.21	10.3	4	23.8	46.7	!	
!	Relais	9.70	26.14	!	0.00	0.0	6	0.0	0.0	!	-0.16	18.7	3	47.5	47.5	!	

6. Interprétez les deux premiers axes (2 points).
7. Les pays Italie et France sont proches dans le premier plan factoriel des disciplines lancers et relais. Que signifie cette position ? (1 point)

Exercice 2 (6 points)

Le but est d'analyser les températures durant l'année dans 35 grandes villes européennes, tableau présenté en dernière page. Les variables considérées sont : janv, fev, mars, avr, mai, juin, juil, aout, sept, oct, nov, dec (température moyenne le mois en question, toutes mesurées en degrés centigrade), moy (moyenne annuelle en degrés centigrades), ampli (amplitude maximale en degrés centigrades), région (cette variable prend les valeurs, est, ouest, nord et sud). Les individus étudiés sont 35 grandes villes d'Europe.

- On réalise une Analyse en Composantes Principales sur ces données (variables centrées et réduites) où les douze mois sont les variables. Pourquoi a-t-on exclu la variable région de l'ACP ? Quelles variables pourraient remplacer région et apporter une information de nature similaire (1 point) ?
- De cette analyse, nous obtenons les valeurs propres des douze composantes principales, $c_j, j=1, \dots, 12$. Leurs valeurs sont les suivantes :

Composantes	1	2	3	4	5	6	7	8	9	10	11	12
Valeurs propres	10,43	1.37	0,12	xxx	0.02	0.009	0.004	0.003	0.002	0.001	0.001	0.000

Définir et déterminer la valeur de l'inertie totale du nuage originel. En déduire la valeur propre associée à la quatrième composante. **(1.5 point)**

- Déterminer combien d'axes vous allez interpréter dans cette analyse. Justifier précisément votre réponse. **(1 point)**
- Le tableau des corrélations des variables avec les deux premières composantes principales est donné ci-dessous.

Corrélations entre les variables et les deux premières composantes principales

variables	janv	fév	mars	avr	mai	juin	juil	aout	sept	oct	nov	dec
Corrélation avec l'axe 1	0.89	0.92	0.96	0.98	0.91	0.88	0.88	0.91	0.99	0.99	0.97	0.91
Corrélation avec l'axe 2	-0.46	-0.39	-0.25	0.06	0.36	0.47	0.46	0.37	0.14	-0.07	-0.24	-0.41

Quel type d'effet mesure-t-on sur le premier axe ? Comment proposez-vous d'interpréter la première composante principale ? **(1,5 point)**

Quelle opposition semble se dessiner sur le second axe ? Comment l'interpréteriez-vous ? **(1 point)**

	Janvier	Février	Mars	Avril	Mai	Juin	Juillet	Août	Septembre	Octobre	Novembre	Décembre	Moyenne	Amplitude	Région
Amsterdam	2.9	2.5	5.7	8.2	12.5	14.8	17.1	17.1	14.5	11.4	7	4.4	9.9	14.6	Ouest
Athènes	9.1	9.7	11.7	15.4	20.1	24.5	27.4	27.2	23.8	19.2	14.6	11	17.8	18.3	Sud
Berlin	-0.2	0.1	4.4	8.2	13.8	16	18.3	18	14.4	10	4.2	1.2	9.1	18.5	Ouest
Bruxelles	3.3	3.3	6.7	8.9	12.8	15.6	17.8	17.8	15	11.1	6.7	4.4	10.3	14.4	Ouest
Budapest	-1.1	0.8	5.5	11.6	17	20.2	22	21.3	16.9	11.3	5.1	0.7	10.9	23.1	Est
Copenhague	-0.4	-0.4	1.3	5.8	11.1	15.4	17.1	16.6	13.3	8.8	4.1	1.3	7.8	17.5	Nord
Dublin	4.8	5	5.9	7.8	10.4	13.3	15	14.6	12.7	9.7	6.7	5.4	9.3	10.2	Nord
Helsinki	-5.8	-6.2	-2.7	3.1	10.2	14	17.2	14.9	9.7	5.2	0.1	-2.3	4.8	23.4	Nord
Kiev	-5.9	-5	-0.3	7.4	14.3	17.8	19.4	18.5	13.7	7.5	1.2	-3.6	7.1	25.3	Est
Cracovie	-3.7	-2	1.9	7.9	13.2	16.9	18.4	17.6	13.7	8.6	2.6	-1.7	7.7	22.1	Est
Lisbonne	10.5	11.3	12.8	14.5	16.7	19.4	21.5	21.9	20.4	17.4	13.7	11.1	15.9	11.4	Sud
Londres	3.4	4.2	5.5	8.3	11.9	15.1	16.9	16.5	14	10.2	6.3	4.4	9.7	13.5	Nord
Madrid	5	6.6	9.4	12.2	16	20.8	24.7	24.3	19.8	13.9	8.7	5.4	13.9	19.7	Sud
Minsk	-6.9	-6.2	-1.9	5.4	12.4	15.9	17.4	16.3	11.6	5.8	0.1	-4.2	5.5	24.3	Est
Moscou	-9.3	-7.6	-2	6	13	16.6	18.3	16.7	11.2	5.1	-1.1	-6	5.1	27.6	Est
Oslo	-4.3	-3.8	-0.6	4.4	10.3	14.9	16.9	15.4	11.1	5.7	0.5	-2.9	5.6	21.2	Nord
Paris	3.7	3.7	7.3	9.7	13.7	16.5	19	18.7	16.1	12.5	7.3	5.2	11.2	15.3	Ouest
Prague	-1.3	0.2	3.6	8.8	14.3	17.6	19.3	18.7	14.9	9.4	3.8	0.3	9.2	20.6	Est
Reykjavik	-0.3	0.1	0.8	2.9	6.5	9.3	11.1	10.6	7.9	4.5	1.7	0.2	4.6	11.4	Nord
Rome	7.1	8.2	10.5	13.7	17.8	21.7	24.4	24.1	20.9	16.5	11.7	8.3	15.4	17.3	Sud
Sarajevo	-1.4	0.8	4.9	9.3	13.8	17	18.9	18.7	15.2	10.5	5.1	0.8	9.4	20.3	Sud
Sofia	-1.7	0.2	4.3	9.7	14.3	17.7	20	19.5	15.8	10.7	5	0.6	9.6	21.7	Est
Stockholm	-3.5	-3.5	-1.3	3.5	9.2	14.6	17.2	16	11.7	6.5	1.7	-1.6	5.8	20.7	Nord
Anvers	3.1	2.9	6.2	8.9	12.9	15.5	17.9	17.6	14.7	11.5	6.8	4.7	10.3	15	Ouest
Barcelone	9.1	10.3	11.8	14.1	17.4	21.2	24.2	24.1	21.7	17.5	13.1	10	16.2	15.1	Sud
Bordeaux	5.6	6.7	9	11.9	15	18.3	20.4	20	17.6	13.5	8.5	6.1	12.7	14.8	Ouest
Edimbourg	2.9	3.6	4.7	7.1	9.9	13	14.7	14.3	12.1	8.7	5.3	3.7	8.3	11.8	Nord
Francfort	0.2	1.8	5.4	9.7	14.3	17.5	19	18.3	14.8	9.8	4.9	1.7	9.8	18.8	Ouest
Genève	0.1	1.9	5.1	9.4	13.8	17.3	19.4	18.5	15	9.8	4.9	1.4	9.7	19.3	Ouest
Gênes	8.7	8.7	11.4	13.8	17.5	21	24.5	24.6	21.8	17.8	12.2	10	16.1	15.9	Sud
Milan	1.1	3.6	8	12.6	17.3	21.3	23.8	22.8	18.9	13.1	6.9	2.6	12.6	22.7	Sud
Palermo	10.5	11.5	13.3	16.9	20.9	23.8	24.5	22.3	22.3	18.4	14.9	12	16.6	14	Sud
Séville	10.7	11.8	14.1	16.1	19.7	23.4	26.7	26.7	24.3	19.4	14.5	11.2	18.2	16	Sud
St. Pétersbov	-8.2	-7.9	-3.7	3.2	10	15.4	18.4	16.9	11.5	5.2	-0.4	-5.3	4.5	26.6	Est
Zurich	-0.7	0.7	4.3	8.5	12.9	16.2	18	17.2	14.1	8.9	3.9	0.3	8.7	18.7	Ouest