

Examen MAT 5201 DATA MINING**Vendredi 27 Novembre 2009****Première Partie : 15 minutes****Enseignant responsable : Frédéric Bertrand**

Remarque important : les questions de ce questionnaire sont posées dans le contexte d'un cours de DATA MINING. Une seule réponse est bonne par question.

1. **Que signifie ACP ?**
 - a) **Analyse en Composantes Principales**
 - b) **Analyse des Correspondances Premières**
 - c) **Analyse des Classes Primaires**
2. **Que signifie GLM ?**
 - a) **Généralisation des Lois Multidimensionnelles**
 - b) **Generalized Linear Model**
 - c) **Gestion Logistique des Modèles**
3. **Que signifie ANOVA ?**
 - a) **Analyse Normalisée et Orientée des Variables Auxiliaires**
 - b) **Association Nationale Orientée des Ventes et des Assurances**
 - c) **ANalysis Of VAriance**
4. **Que signifie CRM ?**
 - a) **Centre de Recherche en Mathématiques**
 - b) **Classification des Relations Maximales**
 - c) **Customer Relationship Management**
5. **Que signifie GRC ?**
 - a) **Groupe de Recherche en Cryptographie**
 - b) **Gestion de la Relation Client**
 - c) **General Regression Classification**
6. **Que signifie CART ?**
 - a) **Classification And Regression Tree**
 - b) **Classement et Analyse dans les Réseaux Téléphoniques**
 - c) **Classification Ascendante des Régressions et des Tests**
7. **Que signifie SVM ?**
 - a) **Simplification des Variables et des Modèles**
 - b) **Segmentation des Vecteurs et des Méthodes**
 - c) **Support Vector Machines**
8. **Le Data Mining est-il utile en CRM ? Donnez un exemple de problématique liée à son utilisation.**
 - a) **Non.**
 - b) **Oui.**

9. Que signifie VPC ?
 - a) Vente Par Correspondance
 - b) Vecteurs Principaux des Composantes
 - c) Valorisation et Partitionnement des Classes
10. Combien de grandes familles de techniques de DATA MINING sont présentées dans ce cours ? Citez-les dans le cas que vous avez choisi.
 - a) Une.
 - b) Deux.
 - c) Trois.
11. Que signifie MANCOVA ?
 - a) Modèle Analytique Normalisé pour une Correspondance Orientée de Variables Automatiques
 - b) Méthode et Analyse Normalisées pour des Classes Orientées vers les Vecteurs Ascendants
 - c) Multiple ANalysis Of COVariance
12. Le logiciel R est
 - a) Libre
 - b) Commercialisé
 - c) Un progiciel
13. Une ACP se réalise
 - a) Sur des variables qualitatives
 - b) Sur des variables quantitatives
 - c) Sur des variables mixtes
14. La commande sous R pour réaliser une ACP est
 - a) ACP()
 - b) Res.pca()
 - c) PCA()
15. La commande sous R pour réaliser une AFC est
 - a) Res.ca()
 - b) CA()
 - c) Assoc()
16. Que signifie CAH ?
 - a) Classification Ascendante Hiérarchique
 - b) Classement Automatique et Homogène
 - c) Composantes Aléatoires et Hiérarchiques

UE de cinquième année : **MAT 5201 - Data Mining**
Enseignants Responsables : F. Bertrand

Chaque réponse devra être justifiée précisément. En annexe sont donnés le journal et la sortie d'un traitement avec le logiciel R.

Exercice :

On dispose des précipitations mensuelles (en mm) (moyennes mensuelles calculées sur 30 ans) pour 34 villes de France. Les villes choisies recouvrent à peu près uniformément le territoire français. Les données sont fournies dans le tableau 1.

1. Décrire le jeu de données (nombre d'individus, nombre de variables, nature des variables)
2. Que pouvez-vous dire à partir des données centrées-réduites (tableau 3)?
3. On veut effectuer une ACP sur ce jeu de données : quels sont les objectifs d'une telle analyse?
4. Les variables ont été centrées et réduites avant l'analyse. La réduction était-elle indispensable? Justifier.
5. Les tableaux 5, 6, 7, 8 et 9 donnent les PRINCIPAUX résultats de l'ACP sur les variables et les individus. Quelle est l'inertie expliquée par le premier axe de l'ACP? Et par le premier plan?
6. Quelles sont les villes qui contribuent le plus à la construction des deux premiers axes? Que signifie une contribution importante?
7. La figure 4 donne le graphe des individus de l'ACP. La figure 5 donne le graphe des variables. Interpréter les facteurs principaux de l'ACP (à l'aide du graphe des individus et de celui des variables).
8. À partir du cercle de corrélations, que pouvez-vous dire concernant les corrélations suivantes février-mars, février-juin?
9. VRAI ou FAUX? Si FAUX, corriger la phrase proposée.
 - Une ville pluvieuse en juillet est également pluvieuse en octobre.
 - La variable janvier est bien représentée sur l'axe 1.
 - La ville de Vichy a joué le rôle le plus important dans la construction de l'axe 2.
 - La coordonnée d'une variable sur un axe est un indicateur de sa qualité de représentation par l'axe.

	Janvier	Février	Mars	Avril	Mai	Juin	Juillet	Août	Septembre	Octobre	Novembre	Décembre
Ajaccio	78	69	51	39	43	23	10	15	43	81	105	96
Angers	65	50	60	45	50	55	35	60	55	65	80	70
Angoulême	79	68	64	62	70	58	53	66	69	70	79	88
Besançon	94	87	75	74	86	107	80	116	106	78	92	93
Biarritz	128	105	98	102	100	91	69	123	155	152	175	176
Bordeaux	100	84	66	57	64	71	52	65	88	84	99	117
Brest	130	98	89	77	74	60	51	80	95	108	136	159
Caen	65	61	45	44	53	52	45	57	66	75	79	71
Clermont-Fd	28	27	30	41	78	79	48	70	58	43	39	30
Dijon	62	48	51	48	68	79	44	79	74	53	67	61
Embrun	61	55	55	48	47	63	41	65	60	60	81	62
Grenoble	80	79	69	69	83	94	74	96	88	85	90	98
Lille	45	43	38	37	45	57	62	64	53	56	56	56
Limoges	87	75	68	69	72	71	56	73	87	72	82	98
Lyon	53	50	60	54	67	84	55	104	86	73	80	62
Marseille	36	49	40	35	38	33	13	27	65	67	69	61
Montpellier	56	59	69	46	47	41	20	52	78	125	70	73
Nancy	66	58	43	45	62	70	58	76	65	52	59	67
Nantes	83	65	53	48	54	52	42	66	80	77	95	94
Nice	67	83	71	70	39	37	21	38	83	109	158	92
Nîmes	52	53	57	45	50	40	25	40	75	100	83	60
Orléans	57	48	43	46	52	54	47	54	51	54	61	54
Paris	53	48	40	45	53	57	54	61	54	50	58	51
Perpignan	27	52	59	47	49	33	27	28	69	97	70	71
Poitiers	65	58	56	49	55	55	46	59	52	61	78	68
Reims	43	44	42	37	52	53	47	58	54	43	52	50
Rennes	57	50	45	43	46	48	36	57	53	60	73	66
Rouen	65	58	50	44	50	57	49	67	70	72	68	66
St-Quentin	52	50	46	44	52	63	61	69	67	52	63	65
Strasbourg	51	44	42	58	71	88	73	90	61	43	51	47
Toulon	76	86	82	60	49	35	12	31	77	105	117	107
Toulouse	53	50	52	55	65	65	44	43	57	49	58	65
Tours	63	55	52	51	53	58	47	60	60	55	68	65
Vichy	50	45	51	52	84	84	63	86	75	58	58	55

TAB. 1 – Données brutes

Janvier	Février	Mars	Avril	Mai	Juin
Min.: 27.00	Min.: 27.00	Min.: 30.00	Min.: 35.00	Min.: 38.00	Min.: 23.00
1st Qu.: 52.25	1st Qu.: 49.25	1st Qu.: 45.00	1st Qu.: 44.25	1st Qu.: 49.25	1st Qu.: 52.00
Median : 62.50	Median : 55.00	Median : 52.50	Median : 48.00	Median : 53.00	Median : 57.50
Mean : 65.50	Mean : 60.41	Mean : 56.24	Mean : 52.53	Mean : 59.44	Mean : 60.80
3rd Qu.: 77.50	3rd Qu.: 68.75	3rd Qu.: 65.50	3rd Qu.: 57.75	3rd Qu.: 69.50	3rd Qu.: 71.00
Max.: 130.00	Max.: 105.00	Max.: 98.00	Max.: 102.00	Max.: 100.00	Max.: 107.00

Juillet	Août	Septembre	Octobre	Novembre	Décembre
Min.: 10.00	Min.: 15.00	Min.: 43.00	Min.: 43.00	Min.: 39.00	Min.: 30.00
1st Qu.: 37.25	1st Qu.: 54.75	1st Qu.: 57.25	1st Qu.: 54.25	1st Qu.: 61.50	1st Qu.: 61.00
Median : 47.00	Median : 64.50	Median : 68.00	Median : 68.50	Median : 75.50	Median : 66.50
Mean : 45.88	Mean : 64.56	Mean : 71.44	Mean : 73.06	Mean : 80.85	Mean : 76.88
3rd Qu.: 55.75	3rd Qu.: 75.25	3rd Qu.: 79.50	3rd Qu.: 83.25	3rd Qu.: 88.25	3rd Qu.: 92.75
Max.: 80.00	Max.: 123.00	Max.: 155.00	Max.: 152.00	Max.: 175.00	Max.: 176.00

	Janvier	Février	Mars	Avril	Mai	Juin	Juillet	Août	Septembre	Octobre	Novembre	Décembre
Janvier	1.00	0.90	0.79	0.78	0.48	0.28	0.27	0.43	0.69	0.56	0.76	0.92
Février	0.90	1.00	0.89	0.82	0.34	0.11	0.07	0.25	0.75	0.75	0.88	0.94
Mars	0.79	0.89	1.00	0.85	0.41	0.15	0.03	0.30	0.80	0.83	0.83	0.88
Avril	0.78	0.82	0.85	1.00	0.70	0.49	0.40	0.58	0.86	0.62	0.75	0.80
Mai	0.48	0.34	0.41	0.70	1.00	0.86	0.73	0.82	0.65	0.14	0.18	0.39
Juin	0.28	0.11	0.15	0.49	0.86	1.00	0.89	0.93	0.48	-0.16	-0.06	0.10
Juillet	0.27	0.07	0.03	0.40	0.73	0.89	1.00	0.87	0.34	-0.26	-0.14	0.07
Août	0.43	0.25	0.30	0.58	0.82	0.93	0.87	1.00	0.62	0.04	0.12	0.26
Septembre	0.69	0.75	0.80	0.86	0.65	0.48	0.34	0.62	1.00	0.73	0.70	0.76
Octobre	0.56	0.75	0.83	0.62	0.14	-0.16	-0.26	0.04	0.73	1.00	0.82	0.77
Novembre	0.76	0.88	0.83	0.75	0.18	-0.06	-0.14	0.12	0.70	0.82	1.00	0.87
Décembre	0.92	0.94	0.88	0.80	0.39	0.10	0.07	0.26	0.76	0.77	0.87	1.00

TAB. 2 – Statistiques descriptives et corrélations des données brutes

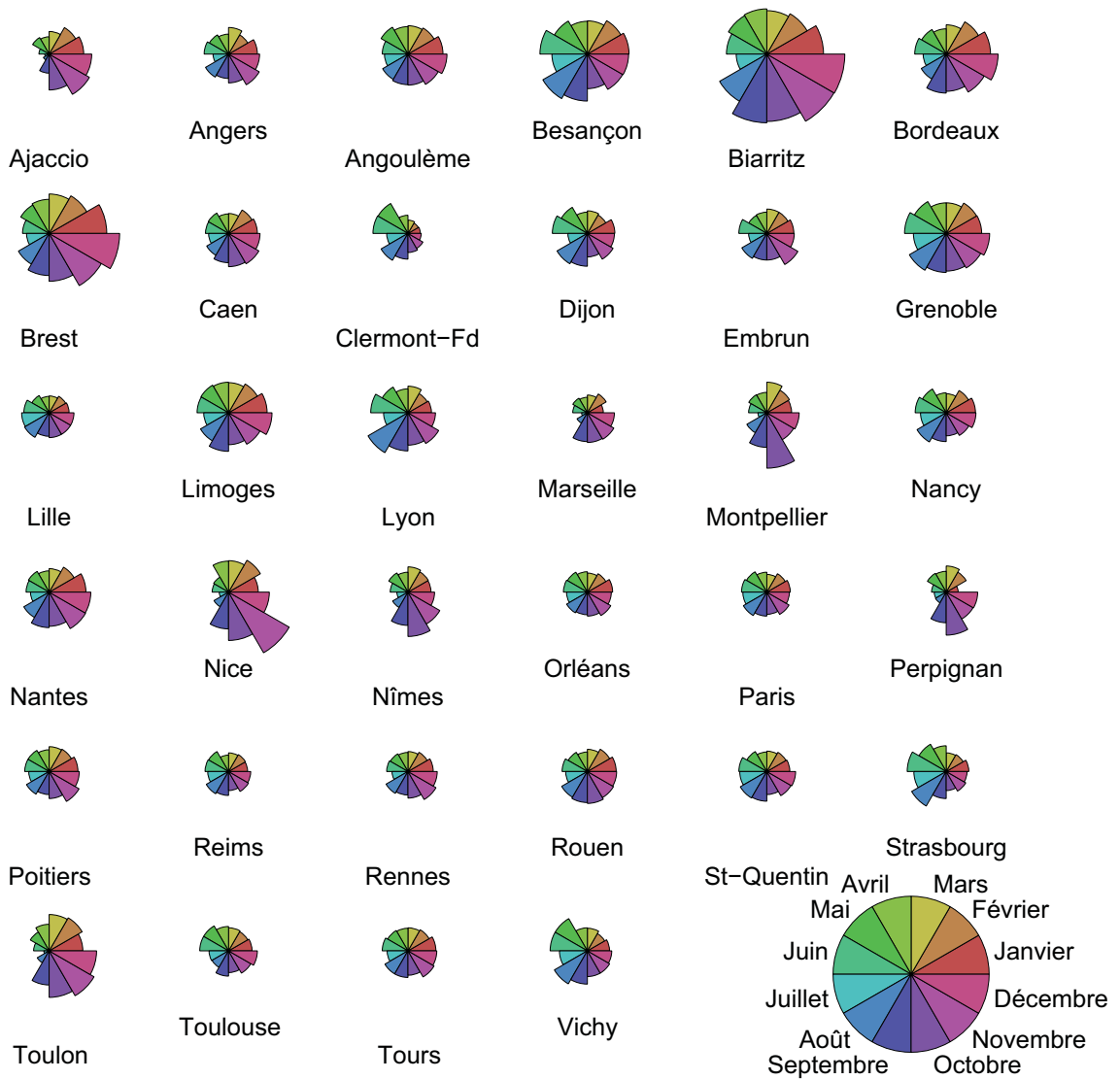


FIG. 1 – *Pluviométrie par ville, données brutes*

	Janvier	Février	Mars	Avril	Mai	Juin	Juillet	Août	Septembre	Octobre	Novembre	Décembre
Ajaccio	0.54	0.49	-0.35	-0.97	-1.10	-1.94	-2.02	-2.06	-1.38	0.31	0.82	0.64
Angers	-0.02	-0.60	0.25	-0.54	-0.63	-0.30	-0.61	-0.19	-0.79	-0.32	-0.03	-0.23
Angoulême	0.58	0.43	0.51	0.68	0.71	-0.14	0.40	0.06	-0.12	-0.12	-0.06	0.37
Besançon	1.23	1.52	1.24	1.54	1.78	2.37	1.92	2.14	1.67	0.20	0.38	0.54
Biarritz	2.70	2.55	2.76	3.55	2.72	1.55	1.30	2.43	4.04	3.13	3.21	3.32
Bordeaux	1.49	1.35	0.65	0.32	0.31	0.52	0.34	0.02	0.80	0.43	0.62	1.34
Brest	2.78	2.15	2.17	1.76	0.98	-0.04	0.29	0.64	1.14	1.38	1.88	2.75
Caen	-0.02	0.03	-0.74	-0.61	-0.43	-0.45	-0.05	-0.31	-0.26	0.08	-0.06	-0.20
Clermont-Fd	-1.62	-1.91	-1.74	-0.83	1.25	0.93	0.12	0.23	-0.65	-1.19	-1.43	-1.57
Dijon	-0.15	-0.71	-0.35	-0.33	0.57	0.93	-0.11	0.60	0.12	-0.79	-0.47	-0.53
Embrun	-0.19	-0.31	-0.08	-0.33	-0.83	0.11	-0.28	0.02	-0.55	-0.52	0.01	-0.50
Grenoble	0.63	1.06	0.84	1.18	1.58	1.70	1.58	1.31	0.80	0.47	0.31	0.71
Lille	-0.88	-1.00	-1.21	-1.12	-0.97	-0.19	0.91	-0.02	-0.89	-0.68	-0.85	-0.70
Limoges	0.93	0.83	0.78	1.18	0.84	0.52	0.57	0.35	0.75	-0.04	0.04	0.71
Lyon	-0.54	-0.60	0.25	0.11	0.51	1.19	0.51	1.64	0.70	-0.00	-0.03	-0.50
Marseille	-1.27	-0.65	-1.07	-1.26	-1.44	-1.43	-1.85	-1.56	-0.31	-0.24	-0.40	-0.53
Montpellier	-0.41	-0.08	0.84	-0.47	-0.83	-1.02	-1.46	-0.52	0.32	2.06	-0.37	-0.13
Nancy	0.02	-0.14	-0.88	-0.54	0.17	0.47	0.68	0.48	-0.31	-0.83	-0.75	-0.33
Nantes	0.75	0.26	-0.21	-0.33	-0.37	-0.45	-0.22	0.06	0.41	0.16	0.48	0.57
Nice	0.06	1.29	0.98	1.26	-1.37	-1.22	-1.40	-1.11	0.56	1.42	2.63	0.51
Nîmes	-0.58	-0.42	0.05	-0.54	-0.63	-1.07	-1.18	-1.02	0.17	1.07	0.07	-0.57
Orléans	-0.37	-0.71	-0.88	-0.47	-0.50	-0.35	0.06	-0.44	-0.99	-0.75	-0.68	-0.77
Paris	-0.54	-0.71	-1.07	-0.54	-0.43	-0.19	0.46	-0.15	-0.84	-0.91	-0.78	-0.87
Perpignan	-1.66	-0.48	0.18	-0.40	-0.70	-1.43	-1.06	-1.52	-0.12	0.95	-0.37	-0.20
Poitiers	-0.02	-0.14	-0.02	-0.25	-0.30	-0.30	0.01	-0.23	-0.94	-0.48	-0.10	-0.30
Reims	-0.97	-0.94	-0.94	-1.12	-0.50	-0.40	0.06	-0.27	-0.84	-1.19	-0.98	-0.90
Rennes	-0.37	-0.60	-0.74	-0.68	-0.90	-0.66	-0.56	-0.31	-0.89	-0.52	-0.27	-0.36
Rouen	-0.02	-0.14	-0.41	-0.61	-0.63	-0.19	0.18	0.10	-0.07	-0.04	-0.44	-0.36
St-Quentin	-0.58	-0.60	-0.68	-0.61	-0.50	0.11	0.85	0.18	-0.21	-0.83	-0.61	-0.40
Strasbourg	-0.63	-0.94	-0.94	0.39	0.78	1.40	1.53	1.06	-0.50	-1.19	-1.02	-1.00
Toulon	0.45	1.46	1.70	0.54	-0.70	-1.32	-1.91	-1.40	0.27	1.26	1.23	1.01
Toulouse	-0.54	-0.60	-0.28	0.18	0.37	0.22	-0.11	-0.90	-0.70	-0.95	-0.78	-0.40
Tours	-0.11	-0.31	-0.28	-0.11	-0.43	-0.14	0.06	-0.19	-0.55	-0.72	-0.44	-0.40
Vichy	-0.67	-0.88	-0.35	-0.04	1.65	1.19	0.96	0.89	0.17	-0.60	-0.78	-0.73

Janvier	Février	Mars	Avril	Mai	Juin
Min. :-1.660e+00	Min. :-1.912e+00	Min. :-1.736e+00	Min. :-1.259e+00	Min. :-1.439e+00	Min. :-1.939e+00
1st Qu. :-5.714e-01	1st Qu. :-6.388e-01	1st Qu. :-7.434e-01	1st Qu. :-5.949e-01	1st Qu. :-6.839e-01	1st Qu. :-4.513e-01
Median :-1.294e-01	Median :-3.097e-01	Median :-2.471e-01	Median :-3.254e-01	Median :-4.322e-01	Median :-1.690e-01
Mean : 1.530e-17	Mean :-1.133e-16	Mean : 4.188e-17	Mean :-1.316e-16	Mean : 2.245e-16	Mean :-1.245e-16
3rd Qu. : 5.175e-01	3rd Qu. : 4.772e-01	3rd Qu. : 6.130e-01	3rd Qu. : 3.751e-01	3rd Qu. : 6.750e-01	3rd Qu. : 5.237e-01
Max. : 2.782e+00	Max. : 2.552e+00	Max. : 2.763e+00	Max. : 3.554e+00	Max. : 2.722e+00	Max. : 2.371e+00

Juillet	Août	Septembre	Octobre	Novembre	Décembre
Min. :-2.021e+00	Min. :-2.063e+00	Min. :-1.375e+00	Min. :-1.190e+00	Min. :-1.429e+00	Min. :-1.572e+00
1st Qu. :-4.863e-01	1st Qu. :-4.084e-01	1st Qu. :-6.861e-01	1st Qu. :-7.448e-01	1st Qu. :-6.607e-01	1st Qu. :-5.324e-01
Median : 6.296e-02	Median :-2.449e-03	Median :-1.664e-01	Median :-1.805e-01	Median :-1.827e-01	Median :-3.480e-01
Mean :-2.554e-17	Mean :-1.377e-16	Mean : 1.588e-16	Mean :-1.348e-16	Mean :-2.084e-16	Mean : 2.383e-16
3rd Qu. : 5.559e-01	3rd Qu. : 4.451e-01	3rd Qu. : 3.896e-01	3rd Qu. : 4.035e-01	3rd Qu. : 2.525e-01	3rd Qu. : 5.319e-01
Max. : 1.922e+00	Max. : 2.433e+00	Max. : 4.040e+00	Max. : 3.126e+00	Max. : 3.214e+00	Max. : 3.322e+00

	Janvier	Février	Mars	Avril	Mai	Juin	Juillet	Août	Septembre	Octobre	Novembre	Décembre
Janvier	1.00	0.90	0.79	0.78	0.48	0.28	0.27	0.43	0.69	0.56	0.76	0.92
Février	0.90	1.00	0.89	0.82	0.34	0.11	0.07	0.25	0.75	0.75	0.88	0.94
Mars	0.79	0.89	1.00	0.85	0.41	0.15	0.03	0.30	0.80	0.83	0.83	0.88
Avril	0.78	0.82	0.85	1.00	0.70	0.49	0.40	0.58	0.86	0.62	0.75	0.80
Mai	0.48	0.34	0.41	0.70	1.00	0.86	0.73	0.82	0.65	0.14	0.18	0.39
Juin	0.28	0.11	0.15	0.49	0.86	1.00	0.89	0.93	0.48	-0.16	-0.06	0.10
Juillet	0.27	0.07	0.03	0.40	0.73	0.89	1.00	0.87	0.34	-0.26	-0.14	0.07
Août	0.43	0.25	0.30	0.58	0.82	0.93	0.87	1.00	0.62	0.04	0.12	0.26
Septembre	0.69	0.75	0.80	0.86	0.65	0.48	0.34	0.62	1.00	0.73	0.70	0.76
Octobre	0.56	0.75	0.83	0.62	0.14	-0.16	-0.26	0.04	0.73	1.00	0.82	0.77
Novembre	0.76	0.88	0.83	0.75	0.18	-0.06	-0.14	0.12	0.70	0.82	1.00	0.87
Décembre	0.92	0.94	0.88	0.80	0.39	0.10	0.07	0.26	0.76	0.77	0.87	1.00

TAB. 4 – Statistiques descriptives et corrélations des données centrées-réduites

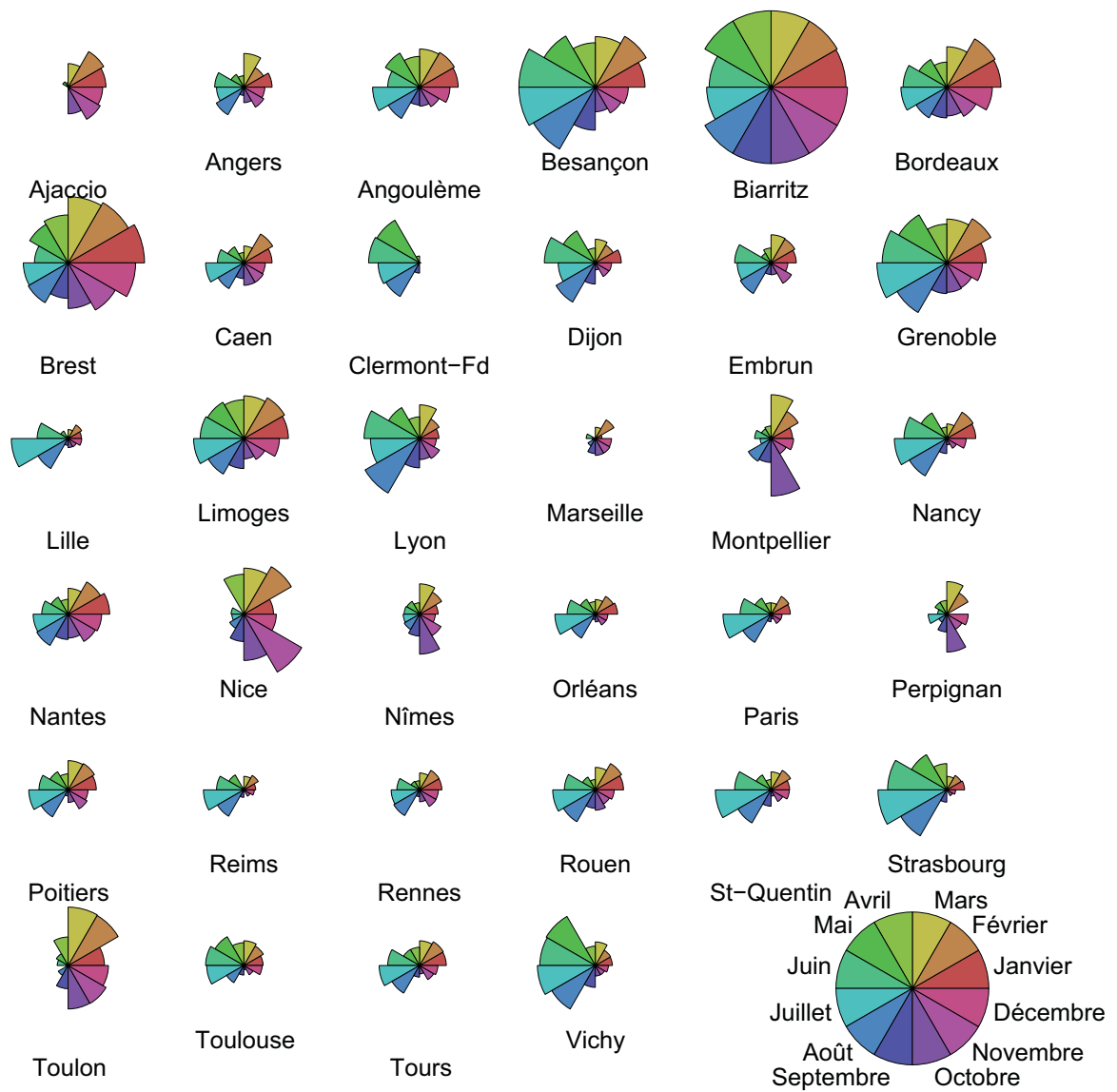


FIG. 2 – Pluviométrie par ville, données centrées-réduites

	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
Janvier	0.89	-0.10	-0.41	0.02	0.14
Février	0.90	-0.32	-0.21	0.00	-0.02
Mars	0.90	-0.29	0.09	-0.10	-0.05
Avril	0.95	0.09	0.04	-0.14	-0.23
Mai	0.65	0.66	0.11	-0.31	0.12
Juin	0.44	0.88	0.05	-0.02	-0.05
Juillet	0.35	0.88	-0.16	0.14	-0.04
Août	0.58	0.77	0.06	0.22	0.02
Septembre	0.92	0.08	0.31	0.12	0.07
Octobre	0.72	-0.55	0.36	0.09	0.12
Novembre	0.81	-0.49	-0.03	0.10	-0.20
Décembre	0.91	-0.31	-0.19	-0.02	0.13

TAB. 5 – *Coordonnées des variables*

	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
Janvier	0.89	-0.10	-0.41	0.02	0.14
Février	0.90	-0.32	-0.21	0.00	-0.02
Mars	0.90	-0.29	0.09	-0.10	-0.05
Avril	0.95	0.09	0.04	-0.14	-0.23
Mai	0.65	0.66	0.11	-0.31	0.12
Juin	0.44	0.88	0.05	-0.02	-0.05
Juillet	0.35	0.88	-0.16	0.14	-0.04
Août	0.58	0.77	0.06	0.22	0.02
Septembre	0.92	0.08	0.31	0.12	0.07
Octobre	0.72	-0.55	0.36	0.09	0.12
Novembre	0.81	-0.49	-0.03	0.10	-0.20
Décembre	0.91	-0.31	-0.19	-0.02	0.13

TAB. 6 – *Corrélations variables - dimensions*

Pourcentage de variance

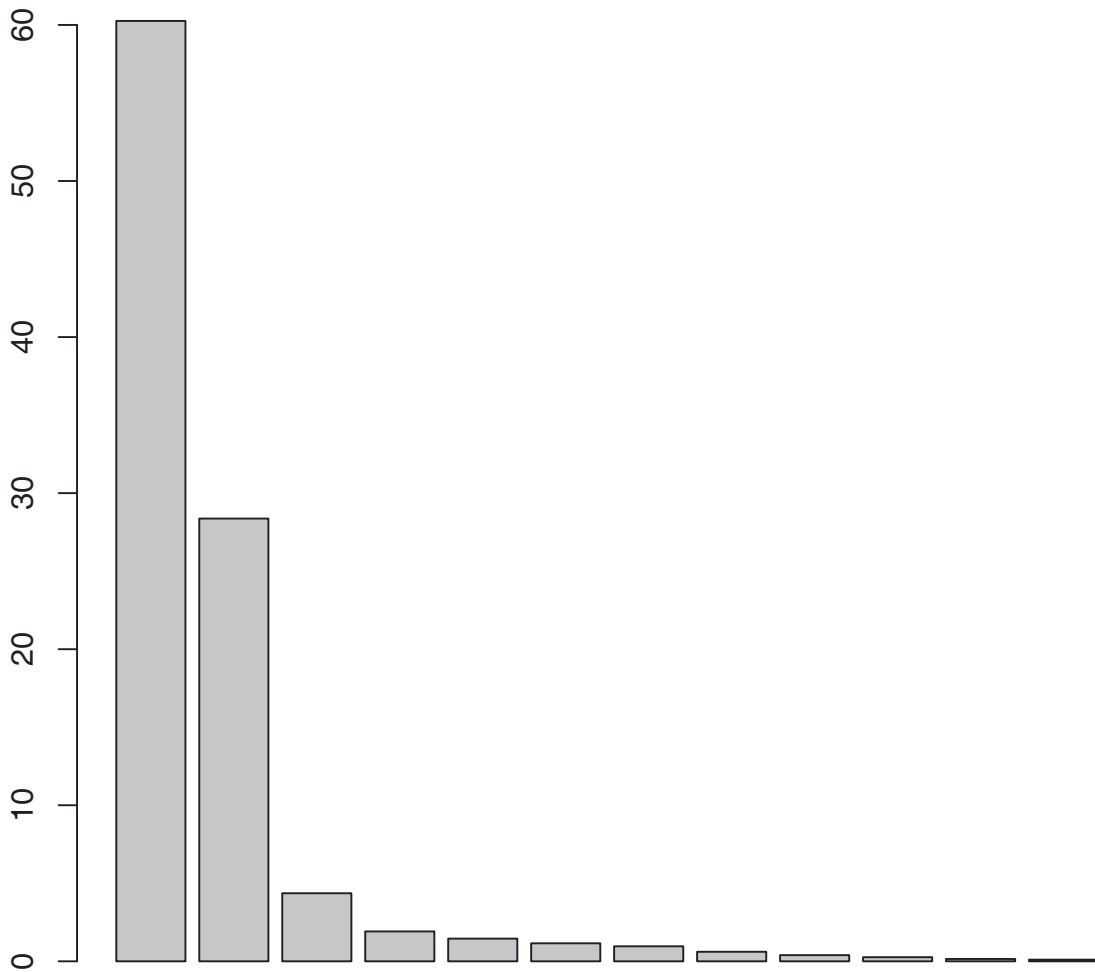


FIG. 3 – Graphe des pourcentages d'inertie des douze axes (sortie **R**)

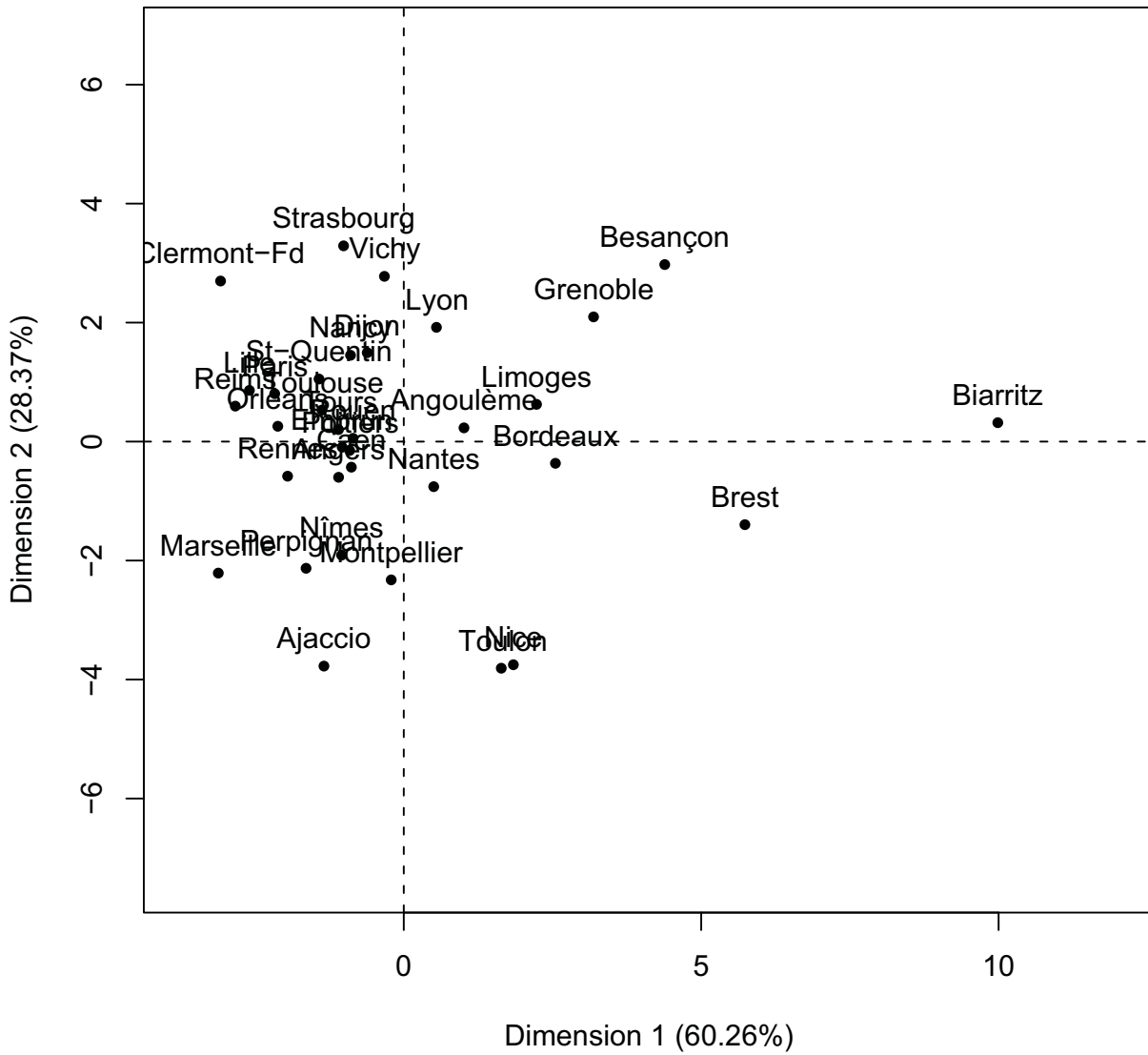


FIG. 4 – Graphe des individus (sortie R)

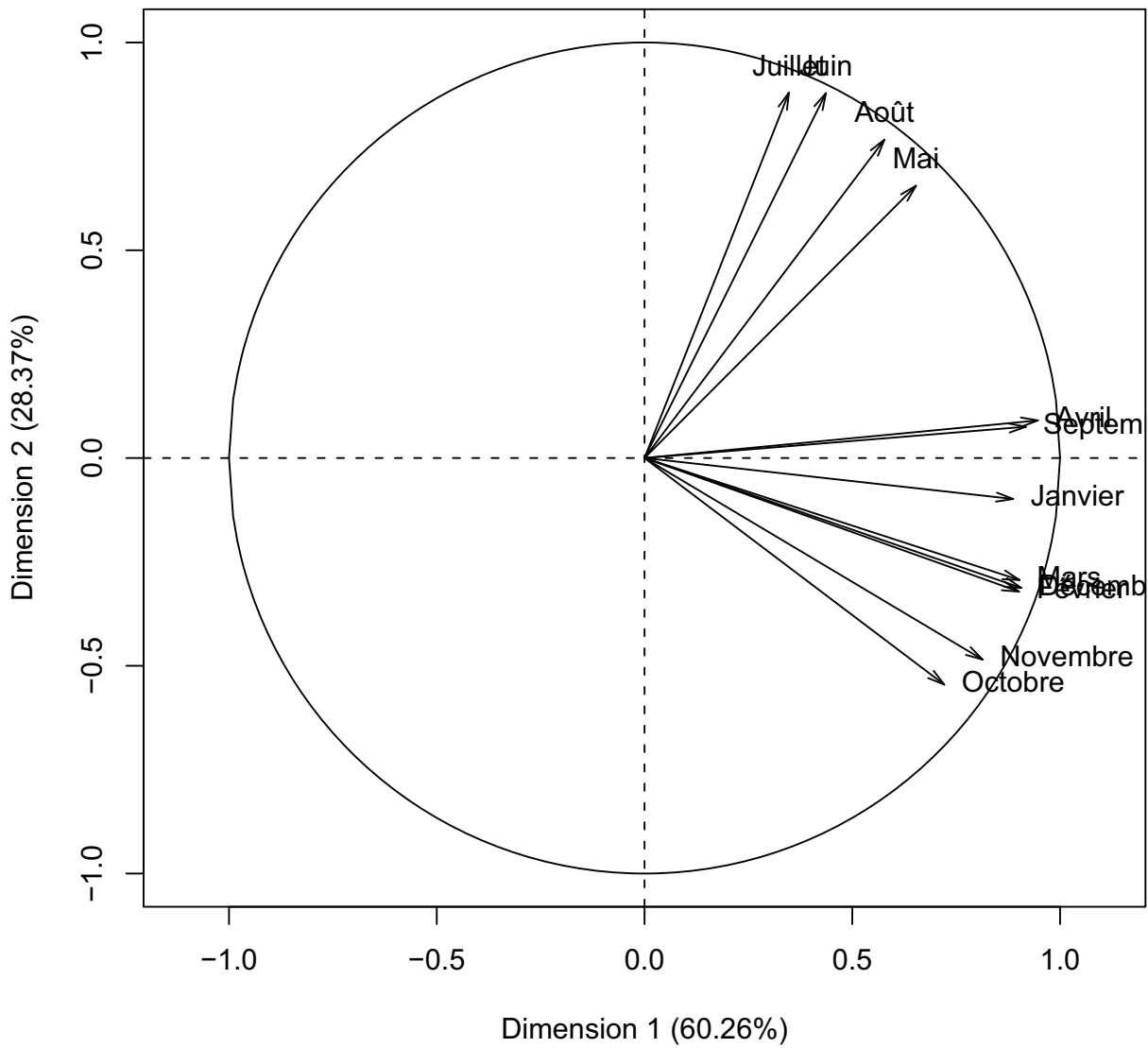


FIG. 5 – Graphe des variables (sortie R)

	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
Janvier	0.79	0.01	0.16	0.00	0.02
Février	0.81	0.10	0.04	0.00	0.00
Mars	0.82	0.09	0.01	0.01	0.00
Avril	0.90	0.01	0.00	0.02	0.05
Mai	0.43	0.43	0.01	0.10	0.01
Juin	0.19	0.77	0.00	0.00	0.00
Juillet	0.12	0.77	0.03	0.02	0.00
Août	0.33	0.59	0.00	0.05	0.00
Septembre	0.84	0.01	0.09	0.01	0.00
Octobre	0.52	0.30	0.13	0.01	0.02
Novembre	0.66	0.24	0.00	0.01	0.04
Décembre	0.82	0.10	0.04	0.00	0.02

TAB. 7 – Cos^2 des variables

	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
Janvier	10.88	0.29	31.40	0.26	11.00
Février	11.24	3.04	8.06	0.01	0.36
Mars	11.27	2.53	1.53	4.41	1.17
Avril	12.40	0.24	0.37	8.57	31.41
Mai	5.90	12.61	2.40	43.03	8.04
Juin	2.64	22.65	0.39	0.13	1.37
Juillet	1.67	22.71	5.17	8.84	0.75
Août	4.61	17.24	0.79	20.33	0.22
Septembre	11.65	0.17	17.98	6.22	2.81
Octobre	7.20	8.73	25.01	3.73	8.78
Novembre	9.16	6.91	0.15	4.27	24.01
Décembre	11.38	2.87	6.75	0.21	10.07

TAB. 8 – Contributions des variables

	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
Ajaccio	-1.34	-3.77	-1.20	-0.50	0.39
Angers	-1.10	-0.60	-0.25	0.00	-0.08
Angoulême	1.01	0.23	-0.54	-0.68	0.01
Besançon	4.39	2.98	-0.04	0.16	-0.20
Biarritz	9.99	0.32	1.07	0.31	0.13
Bordeaux	2.55	-0.36	-0.95	0.10	0.55
Brest	5.74	-1.39	-1.29	-0.26	0.42
Caen	-0.88	-0.43	-0.26	0.43	0.29
Clermont-Fd	-3.08	2.70	1.04	-0.83	0.22
Dijon	-0.61	1.50	0.26	-0.22	0.15
Embrun	-1.04	-0.09	-0.26	0.36	-0.49
Grenoble	3.19	2.10	0.01	-0.20	-0.09
Lille	-2.60	0.86	-0.32	0.95	0.03
Limoges	2.23	0.63	-0.42	-0.55	-0.03
Lyon	0.55	1.92	1.14	0.60	-0.22
Marseille	-3.12	-2.21	0.57	0.08	0.19
Montpellier	-0.21	-2.32	1.66	0.26	0.84
Nancy	-0.90	1.45	-0.61	0.26	0.34
Nantes	0.50	-0.76	-0.50	0.61	0.48
Nice	1.84	-3.75	0.45	0.41	-1.68
Nîmes	-1.05	-1.91	1.20	0.05	0.26
Orléans	-2.12	0.26	-0.46	-0.04	-0.15
Paris	-2.17	0.81	-0.42	0.21	-0.17
Perpignan	-1.64	-2.13	1.49	-0.42	0.05
Poitiers	-0.91	-0.15	-0.62	-0.16	-0.27
Reims	-2.83	0.60	-0.20	0.10	0.05
Rennes	-1.95	-0.58	-0.38	0.28	-0.06
Rouen	-0.85	0.06	-0.12	0.70	0.30
St-Quentin	-1.43	1.05	-0.26	0.63	-0.07
Strasbourg	-1.01	3.29	-0.03	-0.11	-0.56
Toulon	1.64	-3.81	0.13	-0.66	-0.23
Toulouse	-1.37	0.53	-0.18	-1.23	-0.31
Tours	-1.10	0.20	-0.48	-0.04	-0.27
Vichy	-0.33	2.78	0.78	-0.60	0.18

TAB. 9 – *Coordonnées des individus*

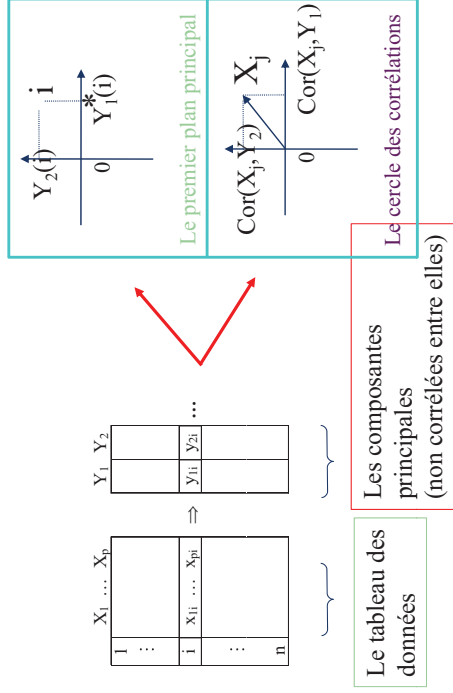
	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
Ajaccio	0.10	0.77	0.08	0.01	0.01
Angers	0.49	0.15	0.03	0.00	0.00
Angoulême	0.47	0.02	0.14	0.21	0.00
Besançon	0.66	0.31	0.00	0.00	0.00
Biarritz	0.98	0.00	0.01	0.00	0.00
Bordeaux	0.78	0.02	0.11	0.00	0.04
Brest	0.89	0.05	0.04	0.00	0.00
Caen	0.49	0.12	0.04	0.12	0.05
Clermont-Fd	0.49	0.38	0.06	0.04	0.00
Dijon	0.11	0.62	0.02	0.01	0.01
Embrun	0.56	0.00	0.04	0.07	0.13
Grenoble	0.67	0.29	0.00	0.00	0.00
Lille	0.75	0.08	0.01	0.10	0.00
Limoges	0.80	0.06	0.03	0.05	0.00
Lyon	0.05	0.59	0.21	0.06	0.01
Marseille	0.61	0.31	0.02	0.00	0.00
Montpellier	0.00	0.54	0.27	0.01	0.07
Nancy	0.22	0.59	0.10	0.02	0.03
Nantes	0.13	0.29	0.12	0.19	0.12
Nice	0.16	0.67	0.01	0.01	0.14
Nîmes	0.17	0.56	0.22	0.00	0.01
Orléans	0.91	0.01	0.04	0.00	0.00
Paris	0.82	0.11	0.03	0.01	0.01
Perpignan	0.26	0.43	0.21	0.02	0.00
Poitiers	0.53	0.01	0.24	0.02	0.05
Reims	0.93	0.04	0.00	0.00	0.00
Rennes	0.84	0.07	0.03	0.02	0.00
Rouen	0.51	0.00	0.01	0.34	0.06
St-Quentin	0.52	0.28	0.02	0.10	0.00
Strasbourg	0.08	0.87	0.00	0.00	0.03
Toulon	0.15	0.80	0.00	0.02	0.00
Toulouse	0.47	0.07	0.01	0.38	0.02
Tours	0.72	0.02	0.14	0.00	0.04
Vichy	0.01	0.86	0.07	0.04	0.00

TAB. 10 – Cos^2 des individus

	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
Ajaccio	0.73	12.29	8.13	3.16	2.59
Angers	0.49	0.31	0.36	0.00	0.12
Angoulême	0.42	0.05	1.65	5.92	0.00
Besançon	7.83	7.66	0.01	0.31	0.68
Biarritz	40.57	0.09	6.40	1.22	0.27
Bordeaux	2.64	0.11	5.02	0.13	5.16
Brest	13.38	1.68	9.30	0.89	2.96
Caen	0.32	0.16	0.39	2.36	1.38
Clermont-Fd	3.86	6.30	6.04	8.87	0.80
Dijon	0.15	1.93	0.39	0.60	0.40
Embrun	0.44	0.01	0.38	1.67	4.12
Grenoble	4.14	3.80	0.00	0.52	0.15
Lille	2.74	0.64	0.56	11.47	0.02
Limoges	2.03	0.34	0.98	3.91	0.01
Lyon	0.12	3.19	7.25	4.64	0.83
Marseille	3.95	4.22	1.82	0.09	0.64
Montpellier	0.02	4.66	15.42	0.89	11.89
Nancy	0.33	1.82	2.12	0.84	1.90
Nantes	0.10	0.49	1.38	4.84	3.87
Nice	1.38	12.15	1.12	2.17	47.63
Nîmes	0.44	3.14	8.09	0.03	1.16
Orléans	1.82	0.06	1.19	0.02	0.39
Paris	1.92	0.56	1.01	0.54	0.49
Perpignan	1.10	3.91	12.40	2.28	0.04
Poitiers	0.34	0.02	2.15	0.31	1.19
Reims	3.26	0.31	0.22	0.14	0.05
Rennes	1.55	0.29	0.80	1.03	0.05
Rouen	0.30	0.00	0.08	6.19	1.53
St-Quentin	0.83	0.96	0.37	5.15	0.09
Strasbourg	0.42	9.36	0.01	0.15	5.28
Toulon	1.09	12.53	0.10	5.56	0.93
Toulouse	0.77	0.24	0.18	19.48	1.65
Tours	0.49	0.04	1.29	0.03	1.19
Vichy	0.04	6.67	3.40	4.63	0.55

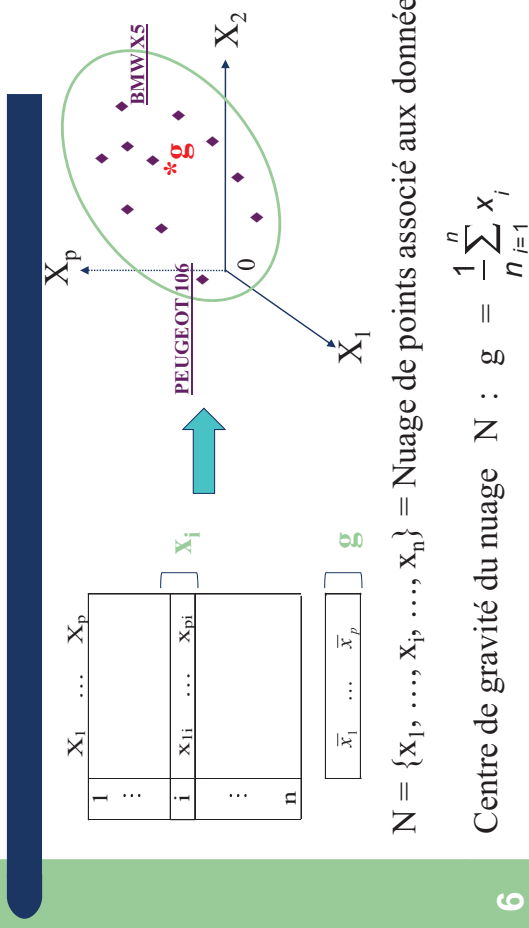
TAB. 11 – *Contributions des individus*

4. Visualisation des données



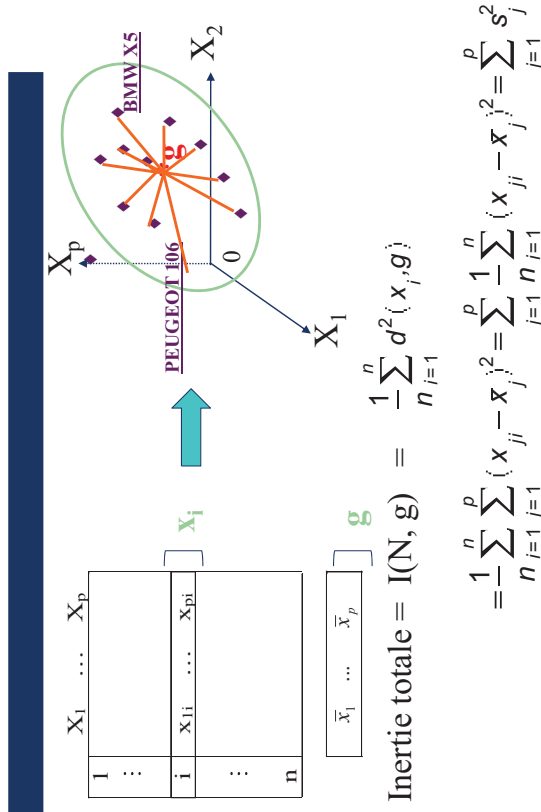
5

5. Le nuage de points associé aux données



6

6. Inertie totale du nuage de points



7

7. Réduction des données

Pour neutraliser le problème des unités on remplace les données d'origine par les données centrées-réduites :

$$X_1^* = \frac{X_1 - \bar{x}_1}{s_1}$$

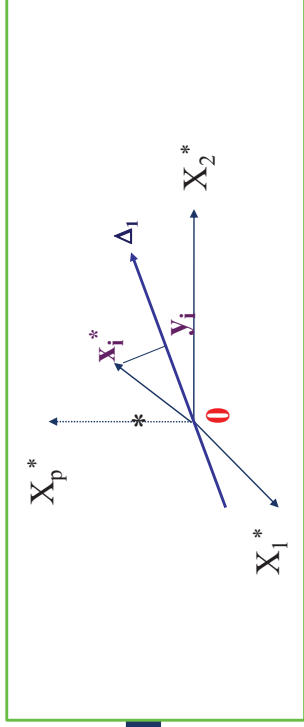
$$\vdots$$

$$X_p^* = \frac{X_p - \bar{x}_p}{s_p}$$

de moyenne 0 et d'écart-type 1.

8

Les objectifs 1 et 2 sont atteints simultanément



De :
on déduit :

$$d^2(x_i, 0) = d^2(y_i, 0) + d^2(x_i, y_i)$$

$$\frac{1}{n} \sum_{i=1}^n d^2(x_i, 0) = \frac{1}{n} \sum_{i=1}^n d^2(y_i, 0) + \frac{1}{n} \sum_{i=1}^n d^2(x_i, y_i)$$

Inertie totale = p

Inertie expliquée par Δ_1

Inertie résiduelle

Maximiser

Minimiser

1er axe principal Δ_1 : Résultats

- L'axe Δ_1 passe par le centre de gravité 0 du nuage de points N^* .
- L'axe Δ_1 est engendré par le vecteur normé u_1 , vecteur propre de la matrice des corrélations R associé à la plus grande valeur propre λ_1 .
- L'inertie du nuage projeté est égal à λ_1 .
- La part d'inertie expliquée par le premier axe principal Δ_1 est égale à λ_1/p .

Résultats SPAD

VALEURS PROPRES
 APERÇU DE LA PRÉCISION DES CALCULS : TRACE AVANT DIAGONALISATION ... 11.0000
 SOMME DES VALEURS PROPRES ... 11.0000

HISTOGRAMME DES 11 PREMIÈRES VALEURS PROPRES

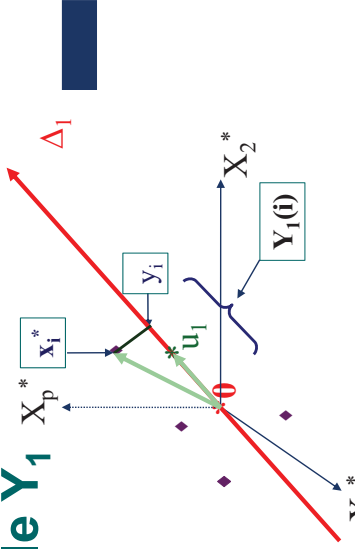
NUMERO	VALEUR PROPRE	POURCENT.	POURCENT. CUMULE
1	6.6969	60.88	60.88
2	2.0236	18.40	79.28
3	0.7451	6.77	86.05
4	0.6926	6.30	92.35
5	0.2839	2.58	94.93
6	0.2013	1.83	96.76
7	0.1300	1.18	97.94
8	0.0893	0.81	98.75
9	0.0757	0.69	99.44
10	0.0385	0.35	99.79
11	0.0230	0.21	100.00

Résultats SPAD

Direction du vecteur propre associé à la plus grande valeur propre :

- 0.86
- 0.91
- 0.54
- 0.87
- 0.90
- 0.21
- 0.46
- 0.94
- 0.94
- 0.76
- 0.81

10. Première composante principale Y_1



Y_1 est une nouvelle variable définie pour chaque individu i par :

$Y_1(i) =$ coordonnée de y_i sur l'axe Δ_1

$=$ produit scalaire entre les vecteurs x_i^* et u_1

$$= \sum_{j=1}^p u_{1j} x_{ij}^* \quad \longrightarrow \quad Y_1 = \sum_{j=1}^p u_{1j} X_j$$

17

18

Interprétation de la première composante principale Y_1

$Y_1 =$	-0.86	PUISS
	-0.91	CYLI
	-0.54	Couplemaxi
	-0.87	LONG
	-0.90	LARG
	-0.21	HAUT
	-0.46	COFFRE
	-0.94	RESE
	-0.94	POIDS
	-0.76	VITE
	-0.81	CONS



Résultats SPAD

COORDONNEES DES INDIVIDUS		COORDONNEES	
AXE 1		P.REL	DISTO
ALF 147 1,9 JTD Distinct	1.10	1.59	0.95
ALF 166 2,5 V6 24V Progr	1.10	5.61	-1.88
ASTWAR DB7 Volante	1.10	42.11	-4.92
AUD A4 3,0 Quattro Pack	1.10	5.09	-1.79
AUD A8 S8 Pack Avus	1.10	26.11	-4.86
AUD TT Roadster 1,8 T225	1.10	5.83	-0.22
AUDIA4 Cabriolet 2,4	1.10	3.11	-1.14
BEN Continental T	1.10	68.44	-7.76
BMW 316i	1.10	1.25	0.23
BMW X5 3,0d Pack Luxe	1.10	14.90	-3.06
BMW Z8	1.10	27.12	-3.68
CAD Seville STS	1.10	21.14	-4.26
CHR Grand Voyager 2,5 CR	1.10	20.40	-2.80

DISTO = $d^2(x_i^*, 0)$

19

Propriétés de la première composante principale Y_1

- Moyenne de $Y_1 = 0$
- Variance de $Y_1 = \frac{1}{n} \sum_{i=1}^n Y_1(i)^2 = \frac{1}{n} \sum_{i=1}^n d^2(y_i, 0) = \lambda_1$
- $\text{Cor}(X_j, Y_1) = \sqrt{\lambda_1} u_{1j}$
- $\frac{1}{p} \sum_{j=1}^p \text{cor}^2(X_j, Y_1) = \frac{\lambda_1}{p}$ est maximum

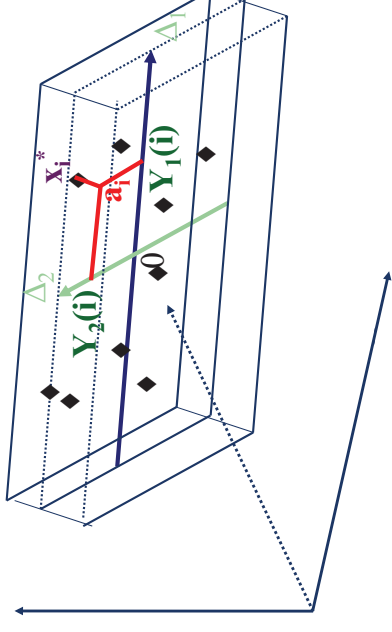
20

Qualité de la première composante principale Y_1

- Inertie totale = 11
- Inertie expliquée par le premier axe principal = $\lambda_1 = 6.69$
- Part d'inertie expliquée par le premier axe principal : $\frac{\lambda_1}{p} = \frac{6.69}{11} = 60.88$
- La première composante principale explique 60.88 % de la variance totale.

21

11. Deuxième axe principal Δ_2



22

2ème axe principal Δ_2 : Résultats

- On recherche le deuxième axe principal Δ_2 orthogonal à Δ_1 et passant le mieux possible au milieu du nuage.
- Il passe par le centre de gravité 0 du nuage de points et est engendré par le vecteur normé u_2 , vecteur propre de la matrice des corrélations R associé à la deuxième plus grande valeur propre λ_2 .
- La deuxième composante principale Y_2 est définie par projection des points sur le deuxième axe principal.
- La deuxième composante principale Y_2 est centrée, de variance λ_2 , et non corrélée à la première composante principale Y_1 .

23

Résultats SPAD

COORDONNEES DES VARIABLES SUR LES AXES 1 A 5
VARIABLES ACTIVES

VARIABLES	COORDONNEES					CORRELATIONS VARIABLE-FACTEUR				
	1	2	3	4	5	1	2	3	4	5
IDEM - LIBELLE COURT										
C2 - RUISS	-0.86	0.43	-0.15	-0.01	0.15	-0.86	0.43	-0.15	-0.01	0.15
C3 - CYLI	-0.91	0.26	-0.20	-0.04	-0.01	-0.91	0.26	-0.20	-0.04	-0.01
C4 - CoupleMaxi	-0.54	0.22	0.53	-0.61	0.02	-0.54	0.22	0.53	-0.61	0.02
C5 - LONG	-0.87	-0.23	0.26	0.23	-0.15	-0.87	-0.23	0.26	0.23	-0.15
C6 - LANG	-0.90	-0.22	0.02	0.06	-0.25	-0.90	-0.22	0.02	0.06	-0.25
C7 - HAUT	-0.21	-0.83	-0.30	-0.35	0.08	-0.21	-0.83	-0.30	-0.35	0.08
C8 - COFFRE	-0.46	-0.73	0.26	0.21	0.33	-0.46	-0.73	0.26	0.21	0.33
C9 - RESE	-0.94	-0.20	-0.04	0.05	-0.12	-0.94	-0.20	-0.04	0.05	-0.12
C10 - POIDS	-0.94	-0.23	-0.09	-0.05	-0.07	-0.94	-0.23	-0.09	-0.05	-0.07
C11 - VITE	-0.76	0.45	0.18	0.26	0.16	-0.76	0.45	0.18	0.26	0.16
C12 - CONS	-0.81	0.29	-0.37	-0.12	0.13	-0.81	0.29	-0.37	-0.12	0.13

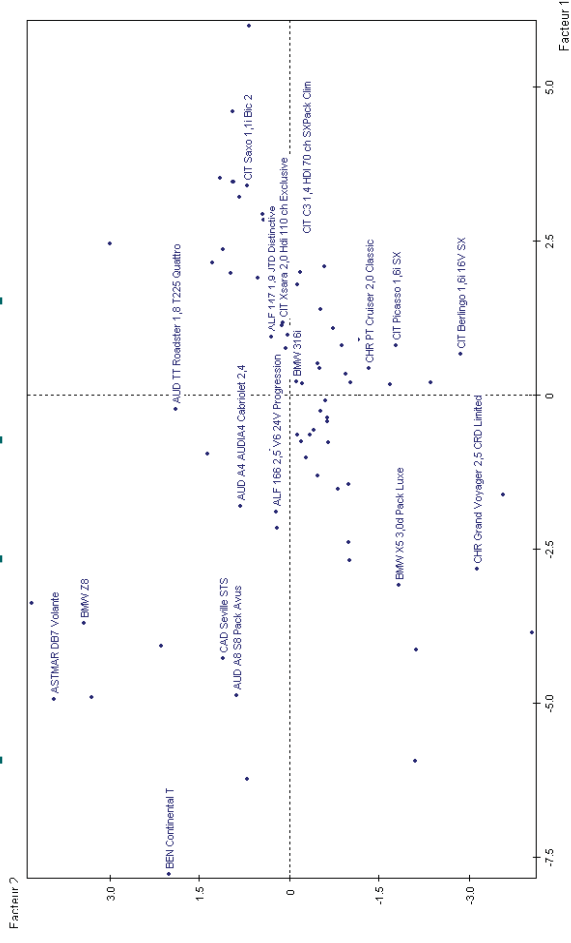
24

Interprétation de la deuxième composante principale Y₂

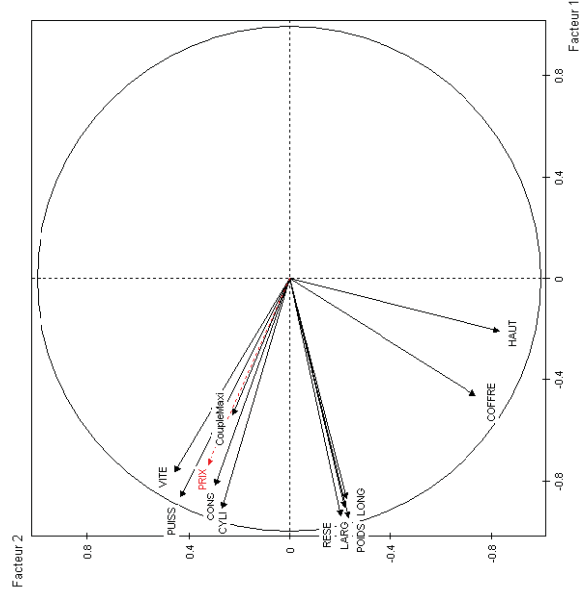
$Y_2 =$
 +0.43 PUISS
 +0.26 CYLI
 +0.22 Couplemaxi
 -0.23 LONG
 -0.22 LARG
 -0.83 HAUT
 -0.73 COFFRE
 -0.20 RESE
 -0.23 POIDS
 +0.45 VITE
 +0.29 CONS

Voiture familiale - 0 + Voiture sportive

12. Exemple Auto 2002 Le premier plan principal



Le cercle des corrélations



13. Qualité globale de l'analyse

Inertie totale = variance totale = p

Part de variance expliquée par la première composante principale = $\frac{\lambda_1}{p}$

Part de variance expliquée par la deuxième composante principale = $\frac{\lambda_2}{p}$

Part de variance expliquée par les deux premières composantes principales = $\frac{\lambda_1 + \lambda_2}{p}$

Et ainsi de suite pour les autres dimensions...