



Université de Poitiers Département de Mathématiques

Statistique descriptive, 1er semestre, année univ. 2009-2010

Fiche 8

Préparation au partiel : sujet du partiel de 2008-09

Exercice 1

Comparaison de groupes

Le taux de PSA (antigène prostatique spécifique) est un critère bien connu permettant de détecter un cancer de la prostate chez l'homme.

Afin de le vérifier, on procède à l'étude suivante : on mesure le taux de PSA dans le sang (exprimé en nanogrammes par millilitre) chez dix hommes sains témoins (groupe T), ainsi que chez dix hommes atteints du cancer de la prostate au stade 1 (groupe C1) et dix hommes atteints du cancer de la prostate au stade 2 (groupe C2).

Les résultats sont les suivants :

groupe T	1.8	2.8	1.4	2.0	0.4	1.4	1.6	1.3	2.7	1.3
groupe C1	2.3	3.3	2.2	3.5	4.1	2.7	2.8	4.0	2.8	2.8
groupe C2	2.7	3.1	2.6	3.9	4.0	2.6	3.3	2.2	4.5	3.8

1. Identifier la population, l'échantillon et les variables en jeu. Donner le type des différentes variables.
2. Pour chaque groupe, calculer la moyenne m , l'écart-type s , le premier quartile Q_1 , la médiane Q_2 , et le troisième quartile Q_3 de l'échantillon donnant le taux de PSA dans le sang. On présentera les résultats dans un tableau.
3. Sur un même graphique, présenter les trois boîtes à moustaches en parallèle représentant la distribution du taux de PSA dans le sang pour chaque groupe (unités à respecter : 2 cm pour 1 ng/ml).
4. Procéder à l'analyse de la variance : donner l'équation de décomposition de la variance, et expliciter les différents termes. Quelle proportion de la variance est expliquée par la répartition en groupe ?
5. Commenter en quelques lignes les différents résultats obtenus : le taux de PSA dans le sang est-il lié au cancer de la prostate ?

Exercice 2

Analyse en composantes principales sur données de températures

On étudie la température moyenne mensuelle au cours des 12 mois de l'année dans 15 villes de France. Les villes retenus dans l'étude sont Bordeaux, Brest, Clermont-Ferrand, Grenoble, Lille, Lyon, Marseille, Montpellier, Nantes, Nice, Paris, Rennes, Strasbourg, Toulouse, Vichy.

La moyenne, l'écart type, le minimum et le maximum de la température mensuelle sont donnés dans le tableau 1.

On procède à une analyse en composante principale normée sur le nuage des individus associés a ces données. Les valeurs propres de la matrice de corrélation sont données dans le tableau 2.

Les résultats de l'ACP sont données de manière graphique : la figure 1 représente le nuage de point projeté sur le premier plan factoriel , la figure 2 représente le cercle des corrélations, sur lequel quatre variables

supplémentaires ont été projetées. Ces variables supplémentaires sont la température moyenne annuelle, l'amplitude de la température, la longitude et la latitude de la ville.

Présenter une interprétation de cette étude en composantes principales. On répondra en particulier aux questions suivantes :

1. Préciser la population, l'échantillon, les variables étudiées.
2. Préciser l'inertie totale du nuage centré réduit des individus, ainsi que l'inertie expliquée par chaque axe principal. Combien d'axes principaux convient-il de conserver ? Quelle est la qualité de représentation pour les deux graphiques présentés ?
3. Pour chacun des deux premiers axes principaux, identifier les individus contribuant fortement à l'axe, ainsi que les variables actives les plus liées à la composante principale associée. Dégager une interprétation pour chacune des composantes principales. (On ne demande pas de prendre en compte les variables supplémentaires dans cette question).
4. Affiner vos commentaires en tenant compte des variables supplémentaires projetées sur le cercle des corrélations associé aux deux premiers facteurs.

TAB. 1 – Résumé des données

Libellé de la variable	Effectif	Moyenne	Ecart-type	Minimum	Maximum
Janvier	15	3,973	1,939	0,400	7,500
Février	15	4,833	1,805	1,500	8,500
Mars	15	8,233	1,477	5,600	10,800
Avril	15	10,980	1,367	8,900	13,300
Mai	15	14,433	1,454	11,600	16,800
Juin	15	17,833	1,733	14,400	20,800
Juillet	15	19,833	2,057	15,600	23,300
Août	15	19,567	1,944	16,000	22,800
Septembre	15	16,987	1,785	14,700	20,300
Octobre	15	12,320	1,768	9,500	16,000
Novembre	15	7,927	1,743	4,900	11,500
Décembre	15	4,847	1,892	1,300	8,200

TAB. 2 – Valeurs propres de l'ACP

Numéro	Valeur propre	Pourcentage	Pourcentage cumulé
1	9,5818	79,85	79,85
2	2,2764	18,97	98,82
3	0,0700	0,58	99,40
4	0,0397	0,33	99,73
5	0,0140	0,12	99,85
6	0,0080	0,07	99,92
7	0,0060	0,05	99,97
8	0,0017	0,01	99,98
9	0,0015	0,01	99,99
10	0,0005	0,00	100,00
11	0,0003	0,00	100,00
12	0,0000	0,00	100,00

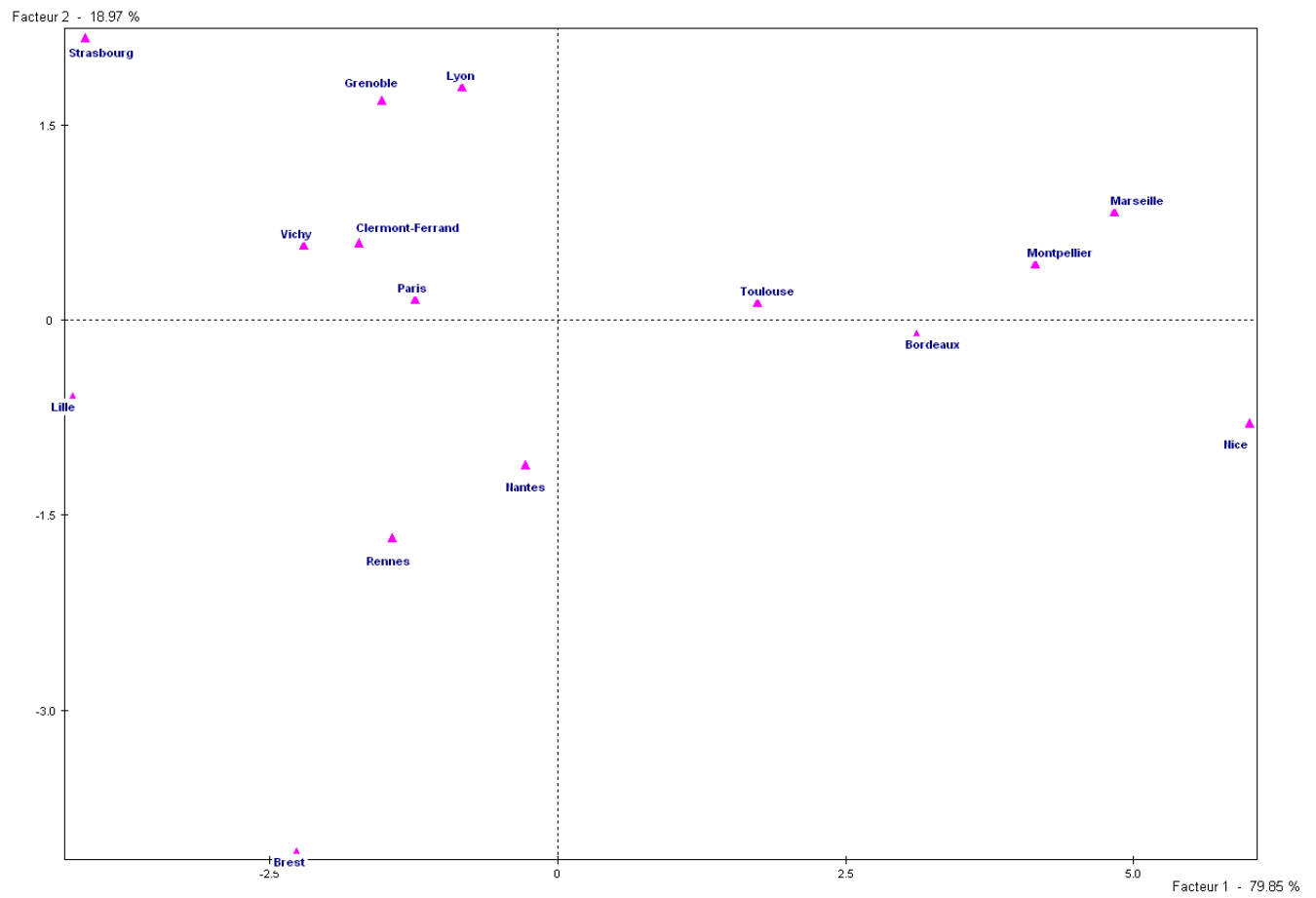


FIG. 1 – Projection du nuage des individus sur le premier plan factoriel.

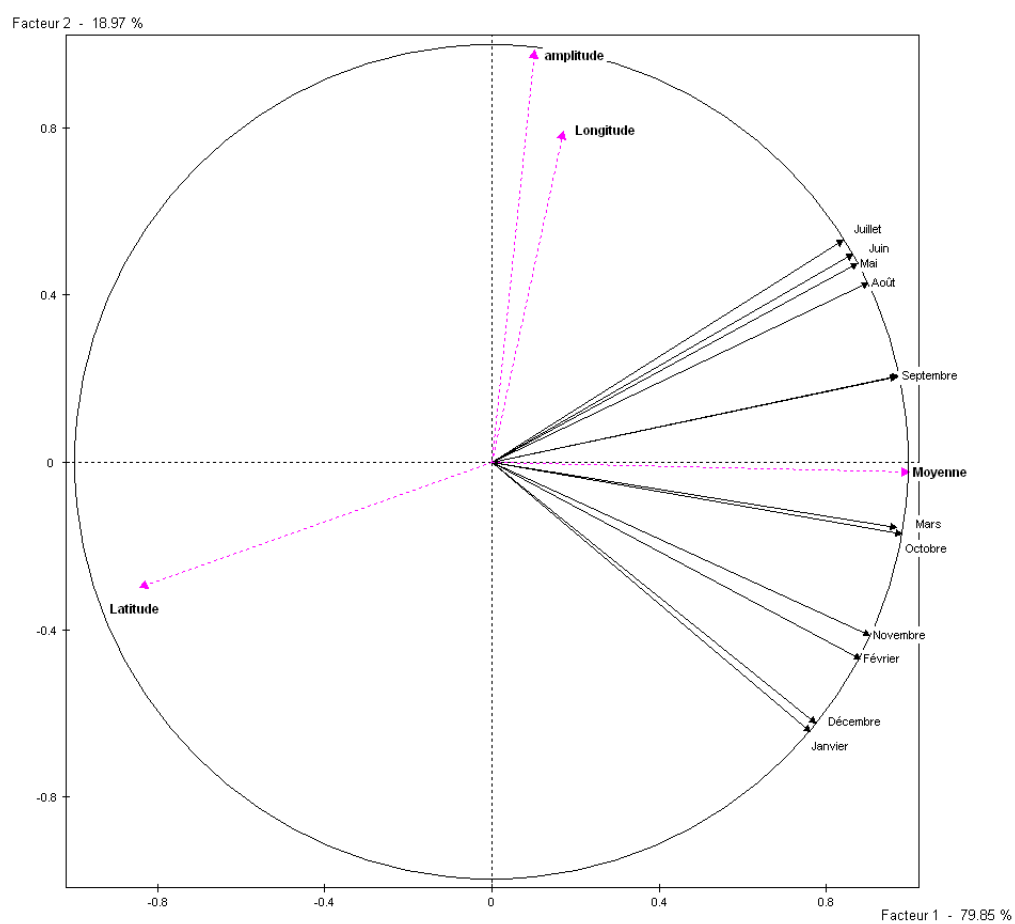


FIG. 2 – Cercle des corrélations avec les 12 variables actives et les 4 variables supplémentaires.