

**Session Principale 2014**

Module	Analyse des données
Auditoire	1ière Année Mastère Ingénierie financière
Enseignant	Mohamed Essaied Hamrita
Durée	Deux heures

*La présentation, la lisibilité, l'orthographe, la qualité de la rédaction et la clarté entreront pour une part importante dans l'appréciation des copies. Aucun document n'est autorisé.*

*Le sujet comporte 4 pages*

## Exercice 1

On dispose d'un fichier `voit` qui contient de données techniques sur 52 véhicules vendus en France (année modèle 1994). Les variables sont : **Puiss** (puissance, en chevaux fiscaux), **Cyl** (cylindrée, en  $\text{cm}^3$ ), **Long** (longueur de la voiture), **Larg** (largeur de la voiture), **Surf** (surface de la voiture) **Poids** (poids total en kg), **Vites** (vitesse maximum en km/h), **DepArr** (temps, en secondes, pour parcourir 1000 m, départ arrêté), **Conso** (consommation moyenne aux 100 km, en litres d'essence ou gazole). Les marques des véhicules sont abrégées **Al** (Alfa-Romeo), **BM** (BMW), **Ci** (Citroen), **Fi** (Fiat), **Pe** (Peugeot) et **Re** (Renault).

- 1) Décrire le jeu de données. La méthode de l'ACP est-elle applicable ? Justifier.
- 2) Est-il préférable d'utiliser l'ACP normée ou l'ACP non normée ? Justifier.
- 3) Dans l'annexe 1, on donne les sorties du logiciel R suite à une ACP normée. Interpréter ces résultats en suivant ce plan :
  - Détermination du nombre d'axes principaux à retenir. Quelle part d'inertie totale sera alors représentée ?
  - Interprétation des axes principaux. (Individus).
  - Interprétation des axes factoriels. (Variables).
  - Interprétation du cercle des corrélations.

## Exercice 2

Le chef d'établissement d'un grand lycée parisien désire comparer les taux de succès au baccalauréat "X" des trois sections générales "Y". Les statistiques de l'année précédente figurent dans le tableau suivant :

	L	ES	S	Total
Réussite	41	59	54	154
Echec	21	36	75	132
Total	62	95	129	286

- 1) Donner la distribution de  $Y$  conditionnellement à  $X$  sous forme de tableau.
- 2) Faire un test d'indépendance pour établir la liaison entre ces caractères.

### Exercice 3

Une entreprise veut choisir un nom pour un nouveau produit, qui doit paraître de qualité supérieure, prestigieuse, luxueuse, convenant à un public masculin, connaisseur, raffiné, distingué, de niveau socio-économique élevé.

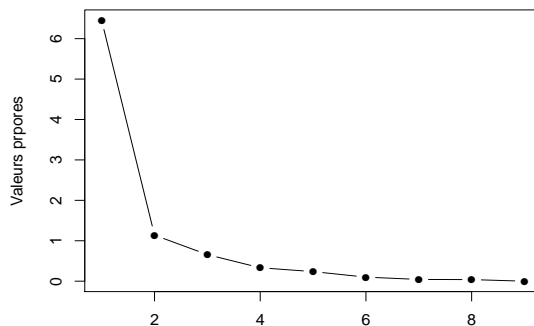
L'entreprise hésite entre 12 marques (cf 1ère ligne du tableau ci-dessous). Pour faire un choix entre ces marques, un échantillon de clients potentiels a été interrogé. Onze attributs leur étaient proposés (cf. 1ère colonne du tableau). Pour chaque attribut, la personne interrogée devait désigner une ou plusieurs marques auxquelles l'attribut se rapportait le mieux.

Les résultats sont fournis dans le listening (Annexe 2).

- 1) Quelle est la méthode d'analyse utilisée ici ? Pourquoi utilise-t-on cette méthode ?
- 2) En examinant le tableau 1 et la figure d'ébouillis des valeurs propres, déterminer le nombre d'axes principaux à retenir. Justifier.
- On désire interpréter le premier axe factoriel en utilisant le deuxième graphique.
- 3) On considère le nuage des attributs. Quels sont les individus dont la contribution est supérieure à la moyenne pour la première variable factorielle ? Pour chacun d'eux, préciser le signe de la coordonnée correspondante. Comment peut-on interpréter cet axe en termes d'opposition entre attributs.
- 4) Même question pour le nuage des marques envisagées (Le premier axe factoriel).
- 5) Mener une étude analogue pour la deuxième variable (Le deuxième axe).

# Annexe 1 : listening exercice 1

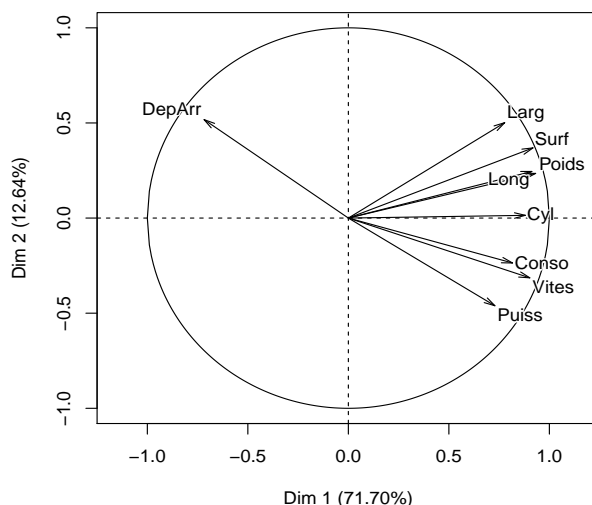
	Variances	% Variances	% Var-Cum
Axe 1	6.453	71.704	71.704
Axe 2	1.138	12.640	84.345
Axe 3	0.657	7.299	91.643
Axe 4	0.332	3.689	95.332
Axe 5	0.237	2.632	97.964
Axe 6	0.100	1.115	99.079
Axe 7	0.043	0.482	99.561
Axe 8	0.039	0.434	99.995
Axe 9	0.000	0.005	100.000



	Coordonnées		Contributions		Qualités	
	Axe 1	Axe 2	Axe 1	Axe 2	Axe 1	Axe 2
Al3315ie	-0.818	-1.624	0.199	4.458	0.142	0.560
Al15520	1.757	-1.345	0.920	3.058	0.587	0.344
Al155TD	0.032	0.771	0.000	1.006	0.001	0.427
Al16425T	2.265	0.332	1.528	0.186	0.621	0.013
BM316i	0.449	-0.077	0.060	0.010	0.174	0.005
BM325TD	1.358	-0.036	0.549	0.002	0.425	0.000
BM518i	2.128	0.005	1.349	0.000	0.788	0.000
BM730i	6.327	-1.429	11.929	3.452	0.925	0.047
CiAX10	-4.112	-1.285	5.039	2.791	0.657	0.064
CiAX14i	-2.878	-2.060	2.468	7.172	0.648	0.332
CiAX14TD	-4.916	0.172	7.202	0.050	0.954	0.001
CiZX11	-2.379	1.018	1.687	1.753	0.732	0.134
CiZX16	-0.876	-0.212	0.229	0.076	0.656	0.038
CiZX19TD	-0.854	0.088	0.217	0.013	0.312	0.003
CiXANTIA16I	-0.021	0.945	0.000	1.511	0.000	0.425
CiXANTIA18I	0.901	0.041	0.242	0.003	0.650	0.001
CiXANTIA20i	1.813	-0.657	0.980	0.729	0.799	0.105
CiXANTIA2TD	0.212	1.180	0.013	2.353	0.026	0.791
CiXM20i	2.856	-0.272	2.430	0.125	0.863	0.008
CiXMV6	4.657	-1.231	6.464	2.563	0.791	0.055
CiXMTD12	1.492	1.508	0.664	3.842	0.382	0.390
FiUNO14ie	-2.657	-1.443	2.103	3.521	0.714	0.211
FiUNOTD	-3.171	-0.904	2.996	1.380	0.845	0.069
FiTIP018ie	0.255	-1.185	0.019	2.374	0.023	0.497
FiTIP0TD	-0.929	0.063	0.257	0.007	0.384	0.002
FiTEMPRA18ie	0.911	-0.833	0.247	1.174	0.416	0.348
FiTEMPRATD	-0.208	0.479	0.013	0.387	0.032	0.168
FiCROMA20	2.026	-0.587	1.223	0.583	0.828	0.070
FiCROMA25TD	1.918	0.292	1.096	0.144	0.734	0.017
Pe106XN	-4.813	0.393	6.902	0.261	0.956	0.006
Pe106XTD	-4.780	0.904	6.809	1.382	0.903	0.032
Pe306XN	-1.847	0.415	1.017	0.291	0.881	0.044
Pe306XTDT	-0.900	-0.040	0.241	0.003	0.428	0.001
Pe405GL	-1.052	0.796	0.330	1.070	0.374	0.214
Pe405SR	0.222	-0.215	0.015	0.078	0.088	0.083
Pe405STD	0.046	0.302	0.001	0.154	0.004	0.158
Pe605SLi	2.659	0.096	2.108	0.015	0.857	0.001
Pe605Sv24	5.643	-2.372	9.489	9.510	0.826	0.146
Pe605SLD	1.023	2.184	0.312	8.063	0.162	0.741
ReTwingo	-3.792	-0.441	4.286	0.329	0.913	0.012
ReCl1012RN	-3.765	0.462	4.224	0.360	0.918	0.014
ReCl1014RN	-2.283	-1.295	1.553	2.836	0.672	0.216
ReCl1019D	-2.702	-0.196	2.176	0.065	0.830	0.004
Re1914RN	-1.104	-0.137	0.363	0.032	0.611	0.009
Re1918RN	-0.308	-0.360	0.028	0.219	0.160	0.219
Re1919dT	-0.594	0.006	0.105	0.000	0.148	0.000
Re21PrimaTS	0.577	-0.244	0.099	0.101	0.145	0.026
Re21PrimaTD	0.269	0.699	0.022	0.826	0.033	0.221
ReSafrane20i	2.455	0.707	1.796	0.845	0.844	0.070
ReSafrane25d	2.613	1.580	2.035	4.222	0.631	0.231
ReEspace22i	3.261	1.557	3.169	4.100	0.514	0.117
ReEspace21dT	1.634	3.484	0.795	20.515	0.147	0.669

	Coordonnées		Contributions		Qualités	
	Axe 1	Axe 2	Axe 1	Axe 2	Axe 1	Axe 2
Puiss	0.729	-0.461	8.236	18.674	0.532	0.212
Cyl	0.880	0.014	12.012	0.018	0.775	0.000
Long	0.913	0.245	12.911	5.278	0.833	0.060
Larg	0.778	0.501	9.372	22.106	0.605	0.251
Surf	0.919	0.369	13.073	11.961	0.844	0.136
Poids	0.932	0.235	13.452	4.846	0.868	0.055
Vites	0.903	-0.314	12.627	8.679	0.815	0.099
DepArr	-0.718	0.517	7.980	23.537	0.515	0.268
Conso	0.817	-0.236	10.337	4.901	0.667	0.056

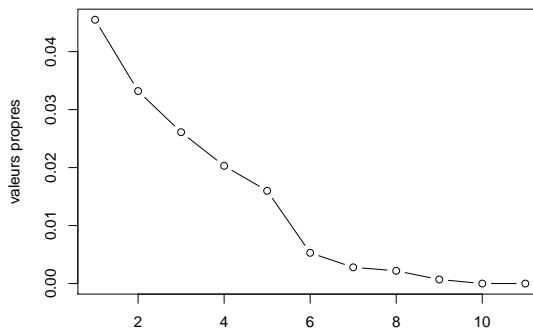
Variables factor map (PCA)



# Annexe 2 : listening exercice 3

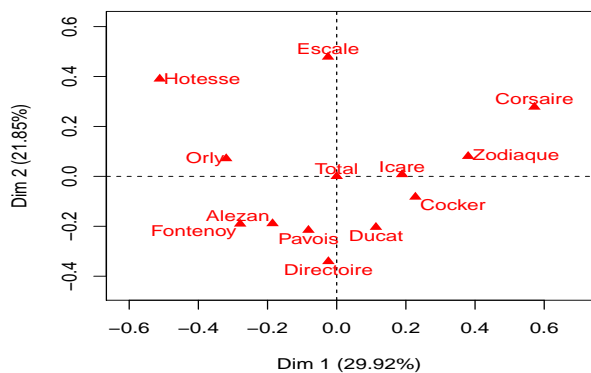
	val.propres	% variance	% cum.var
dim 1	0.0455	29.9210	29.9210
dim 2	0.0332	21.8484	51.7694
dim 3	0.0261	17.1421	68.9115
dim 4	0.0203	13.3400	82.2515
dim 5	0.0160	10.5317	92.7832
dim 6	0.0053	3.4817	96.2649
dim 7	0.0028	1.8232	98.0881
dim 8	0.0022	1.4610	99.5491
dim 9	0.0007	0.4338	99.9829
dim 10	0.0000	0.0171	100.0000
dim 11	0.0000	0.0000	100.0000

Eboulis des valeurs propres



	Poids	Dim 1	Dim 2	Dim 1	Dim 2	Dim 1	Dim 2
Orly	0.034	-0.320	0.071	7.639	0.521	0.419	0.021
Alezan	0.049	-0.186	-0.189	3.698	5.260	0.083	0.086
Corsaire	0.046	0.572	0.278	33.233	10.741	0.572	0.135
Directoire	0.058	-0.024	-0.341	0.073	20.162	0.002	0.439
Ducat	0.036	0.113	-0.204	1.025	4.527	0.094	0.302
Fontenoy	0.037	-0.279	-0.191	6.276	4.013	0.312	0.146
Icare	0.034	0.189	0.007	2.672	0.006	0.310	0.000
Zodiaque	0.031	0.380	0.080	9.838	0.603	0.492	0.022
Pavois	0.034	-0.082	-0.216	0.506	4.825	0.032	0.220
Cocker	0.054	0.228	-0.083	6.108	1.107	0.235	0.031
Escale	0.037	-0.025	0.478	0.051	25.226	0.002	0.754
Hotesse	0.050	-0.512	0.390	28.881	23.011	0.540	0.314
Total	0.500	0.000	0.000	0.000	0.000	NaN	NaN

Projections des marques



Projections des attribus

