COURS 6

Exploitation et traitement des données

Rodolphe Priam, Ingénieur stat & Dr informatique

Partie ACP

- Voir cours ci-joint
- Les diapositives suivantes sont des compléments au cours.

Diagonalisation

- Matrice symmétrique C de taille p x p

- Diagonalisation

  - $C = V S V^T$

    - S matrice diagonal : seule les éléments S[k,k] de la matrice sont non nuls

    - V matrice orthogonale : $V V^T = I$

Diagonalisation – exemple avec Python

```
C = X.corr()
C
```

	CYL	PUISS	LONG	LARG	POIDS	V_MAX
CYL	1.000000	0.796628	0.701462	0.629757	0.788952	0.664934
PUISS	0.796628	1.000000	0.641362	0.520832	0.765293	0.844379
LONG	0.701462	0.641362	1.000000	0.849266	0.868090	0.475928
LARG	0.629757	0.520832	0.849266	1.000000	0.716874	0.472945
POIDS	0.788952	0.765293	0.868090	0.716874	1.000000	0.477596
V_MAX	0.664934	0.844379	0.475928	0.472945	0.477596	1.000000

$$C = V\,S\,V^T$$

Ici $S = \mathrm{diag}(s)$

```
S, V = numpy.linalg.eig(C)
```

```
V @ numpy.diag(S) @ V.T -C
```

	CYL	PUISS	LONG	LARG	POIDS	V_MAX
CYL	-1.665335e-15	-9.992007e-16	-2.220446e-16	-6.661338e-16	-6.661338e-16	-4.440892e-16
PUISS	-9.992007e-16	4.440892e-16	1.110223e-16	1.110223e-16	4.440892e-16	-1.110223e-16
LONG	-2.220446e-16	3.330669e-16	4.440892e-16	2.220446e-16	4.440892e-16	-1.110223e-16
LARG	-6.661338e-16	1.110223e-16	2.220446e-16	-4.440892e-16	1.110223e-16	0.000000e+00
POIDS	-5.551115e-16	4.440892e-16	5.551115e-16	2.220446e-16	-2.220446e-16	6.661338e-16
V_MAX	-5.551115e-16	-1.110223e-16	-1.110223e-16	0.000000e+00	4.440892e-16	-7.771561e-16

```
numpy.round(V@V.T,10)
```

```
array([[ 1., -0.,  0., -0.,  0.,  0.],
       [-0.,  1.,  0., -0., -0.,  0.],
       [ 0.,  0.,  1.,  0., -0., -0.],
       [-0., -0.,  0.,  1.,  0.,  0.],
       [ 0., -0., -0.,  0.,  1.,  0.],
       [ 0.,  0., -0.,  0.,  0.,  1.]])
```

```
S
```

```
array([4.42085806, 0.85606229, 0.37306608, 0.21392209, 0.09280121,
```

Coordonnées des variables en acp via diagonalisat°

```
C = X.corr()
C
```

	CYL	PUISS	LONG	LARG	POIDS	V_MAX
CYL	1.000000	0.796628	0.701462	0.629757	0.788952	0.664934
PUISS	0.796628	1.000000	0.641362	0.520832	0.765293	0.844379
LONG	0.701462	0.641362	1.000000	0.849266	0.868090	0.475928
LARG	0.629757	0.520832	0.849266	1.000000	0.716874	0.472945
POIDS	0.788952	0.765293	0.868090	0.716874	1.000000	0.477596
V_MAX	0.664934	0.844379	0.475928	0.472945	0.477596	1.000000

```
S, V = numpy.linalg.eig(C)
```

```
numpy.round(S, 6)
```

```
array([4.420858, 0.856062, 0.373066, 0.213922, 0.092801, 0.04329 ])
```

```
numpy.round(V, 6)
```

```
array([[-0.424936, -0.124191, -0.353613,  0.807786,  0.15158 , -0.058895],
       [-0.421794, -0.415774, -0.18492 , -0.357792, -0.293735, -0.633033],
       [-0.42146 ,  0.411818,  0.067634, -0.279752,  0.730569, -0.190292],
       [-0.386922,  0.446087,  0.604868,  0.211569, -0.47819 , -0.109566],
       [-0.430512,  0.242676, -0.484396, -0.301711, -0.304558,  0.580812],
       [-0.358944, -0.619863,  0.485472, -0.073574,  0.188655,  0.458522]])
```

```
numpy.round(numpy.sqrt(S) * V, 6)
```

```
array([[-0.893464, -0.114906, -0.215983,  0.373615,  0.046176, -0.012254],
       [-0.886858, -0.384689, -0.112948, -0.165485, -0.089481, -0.131711],
       [-0.886155,  0.381029,  0.04131 , -0.12939 ,  0.222555, -0.039593],
       [-0.813536,  0.412736,  0.369448,  0.097854, -0.145672, -0.022797],
       [-0.905187,  0.224532, -0.295865, -0.139547, -0.092779,  0.120846],
       [-0.75471 , -0.573519,  0.296522, -0.034029,  0.057471,  0.095401]])
```

F1 F2

$$r_{x_j}(F_k) = \sqrt{\lambda_k} a_{jk}$$

Ici a_{jk} composante de V

Calcul de l'ACP par la SVD (1/2)

- Le principe de la SVD est de remplacer deux diagonalisation par une décomposit°

X

	CYL	PUISS	LONG	LARG	POIDS	V_MAX
Alfasud_TI	1350	79	393	161	870	165
Audi_100	1588	85	468	177	1110	160
Simca_1300	1294	68	424	168	1050	152
Citroen_GS_Club	1222	59	412	161	930	151
Fiat_132	1585	98	439	164	1105	165
Lancia_Beta	1297	82	429	169	1080	160
Peugeot_504	1796	79	449	169	1160	154
Renault_16_TL	1565	55	424	163	1010	140
Renault_30	2664	128	452	173	1320	180
Toyota_Corolla	1166	55	399	157	815	140
Alfetta_1.66	1570	109	428	162	1060	175
Princess_1800	1798	82	445	172	1160	158
Datsun_200L	1998	115	469	169	1370	160
Taunus_2000	1993	98	438	170	1080	167
Rancho	1442	80	431	166	1129	144
Mazda_9295	1769	83	440	165	1095	165
Opel_Rekord	1979	100	459	173	1120	173
Lada_1300	1294	68	404	161	955	140

La matrice centrée-réduite $Z = (z_{ij})$ permet d'écrire la matrice de corrélation par un simple produit de matrices (multiplié par un facteur)

$$z_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j} \quad \begin{cases} \bar{x}_j = \frac{1}{n}\sum_{i=1}^{n} x_{ij} \\ s_j^2 = \frac{1}{n}\sum_{i=1}^{n}(x_{ij} - \bar{x}_j)^2 \end{cases}$$

```
((Z.T @ Z)/18 - C)
```
= 0

	CYL	PUISS	LONG	LARG	POIDS	V_MAX
CYL	-2.220446e-16	2.220446e-16	3.330669e-16	5.551115e-16	0.000000e+00	-6.661338e-16
PUISS	2.220446e-16	-2.220446e-16	4.440892e-16	7.771561e-16	1.110223e-16	-5.551115e-16
LONG	3.330669e-16	4.440892e-16	2.220446e-16	2.220446e-16	2.220446e-16	-4.440892e-16
LARG	5.551115e-16	7.771561e-16	2.220446e-16	6.661338e-16	6.661338e-16	5.551115e-17
POIDS	0.000000e+00	1.110223e-16	2.220446e-16	6.661338e-16	2.220446e-16	-3.330669e-16
V_MAX	-6.661338e-16	-5.551115e-16	-4.440892e-16	5.551115e-17	-3.330669e-16	-2.220446e-16

Calcul de l'ACP par la SVD (2/2)

- Le principe de la SVD est de remplacer 2 diagonalisations par 1 décomposition

X

	CYL	PUISS	LONG	LARG	POIDS	V_MAX
Alfasud_TI	1350	79	393	161	870	165
Audi_100	1588	85	468	177	1110	160
Simca_1300	1294	68	424	168	1050	152
Citroen_GS_Club	1222	59	412	161	930	151
Fiat_132	1585	98	439	164	1105	165
Lancia_Beta	1297	82	429	169	1080	160
Peugeot_504	1796	79	449	169	1160	154
Renault_16_TL	1565	55	424	163	1010	140
Renault_30	2664	128	452	173	1320	180
Toyota_Corolla	1166	55	399	157	815	140
Alfetta_1.66	1570	109	428	162	1060	175
Princess_1800	1798	82	445	172	1160	158
Datsun_200L	1998	115	469	169	1370	160
Taunus_2000	1993	98	438	170	1080	167
Rancho	1442	80	431	166	1129	144
Mazda_9295	1769	83	440	165	1095	165
Opel_Rekord	1979	100	459	173	1120	173
Lada_1300	1294	68	404	161	955	140

$C = Z^T Z / n$

$Z = UWV^T$

$\Rightarrow$ $C = V S V^T$ avec $S = W^2/n$

On peut écrire la svd de Z par : $Z = U W V^T$ avec U et V deux matrices orthogonales
W matrice diagonale

D'où $C = Z^T Z/n = (U W V^T)^T (U W V^T)/n = VW U^T UW V^T/n = V (W^2/n) V^T$.

Donc il est retrouvé la diagonalisation de la matrice C
à partir de la decomposition en valeurs singulières de Z.

Ce résultat permet de passer des projections des lignes à celles des colonnes en acp !

Questions ?

Analyse en Composantes Principales (ACP)

Principes et pratique de l'ACP

Ricco RAKOTOMALALA

Université Lumière Lyon 2

PLAN

Position du problème

Construire un nouveau système de représentation

(composantes principales, axes factoriels, facteurs : combinaisons linéaires des variables originelles)

qui permet synthétiser l'information

Variables « actives » quantitatives c.-à-d. seront utilisées pour la construction des facteurs

$j : 1, \ldots, p$

Les données « autos »

(Saporta, 2006 ; page 428)

$i : 1, \ldots, n$

Individus actifs

Modele	CYL	PUISS	LONG	LARG	POIDS	V.MAX
Alfasud TI	1350	79	393	161	870	165
Audi 100	1588	85	468	177	1110	160
Simca 1300	1294	68	424	168	1050	152
Citroen GS Club	1222	59	412	161	930	151
Fiat 132	1585	98	439	164	1105	165
Lancia Beta	1297	82	429	169	1080	160
Peugeot 504	1796	79	449	169	1160	154
Renault 16 TL	1565	55	424	163	1010	140
Renault 30	2664	128	452	173	1320	180
Toyota Corolla	1166	55	399	157	815	140
Alfetta 1.66	1570	109	428	162	1060	175
Princess 1800	1798	82	445	172	1160	158
Datsun 200L	1998	115	469	169	1370	160
Taunus 2000	1993	98	438	170	1080	167
Rancho	1442	80	431	166	1129	144
Mazda 9295	1769	83	440	165	1095	165
Opel Rekord	1979	100	459	173	1120	173
Lada 1300	1294	68	404	161	955	140

x_{ij}

Questions :

(1) Quelles sont les véhicules qui se ressemblent ? (proximité entre les individus)

(2) Sur quelles variables sont fondées les ressemblances / dissemblances

(3) Quelles sont les relations entre les variables

Position du problème (1)

Analyse des proximités entre les individus

Que voit-on dans ce graphique ?

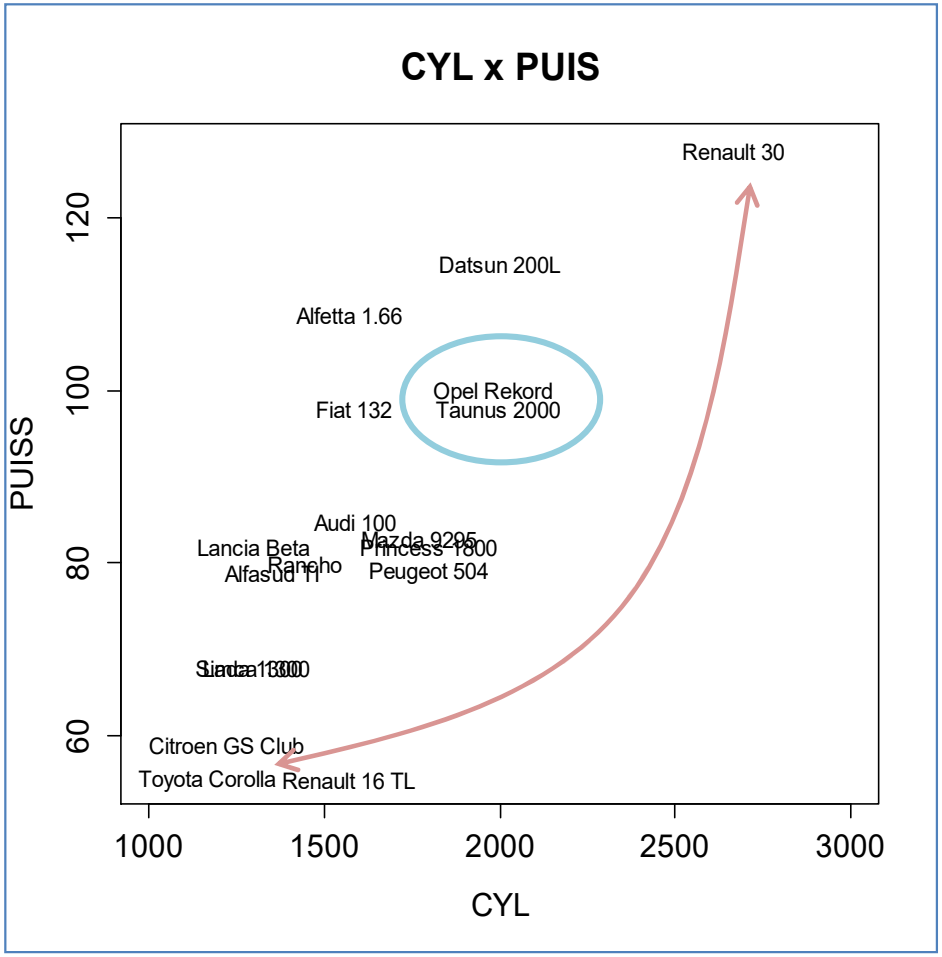
1. Les variables CYL et PUISS sont liées.
2. « Opel Reckord » et « Taunus 2000 (Ford) » ont le même profil (caractéristiques)
3. « Renault 30 » et « Toyota Corolla » ont des profils opposés…

Un graphique ne fait que révéler des informations présentes dans le tableau de données !

Modele	CYL	PUISS
Toyota Corolla	**1166**	**55**
Citroen GS Club	1222	59
Simca 1300	1294	68
Lada 1300	1294	68
Lancia Beta	1297	82
Alfasud TI	1350	79
Rancho	1442	80
Renault 16 TL	1565	55
Alfetta 1.66	1570	109
Fiat 132	1585	98
Audi 100	1588	85
Mazda 9295	1769	83
Peugeot 504	1796	79
Princess 1800	1798	82
Opel Rekord	**1979**	**100**
Taunus 2000	**1993**	**98**
Datsun 200L	1998	115
Renault 30	**2664**	**128**

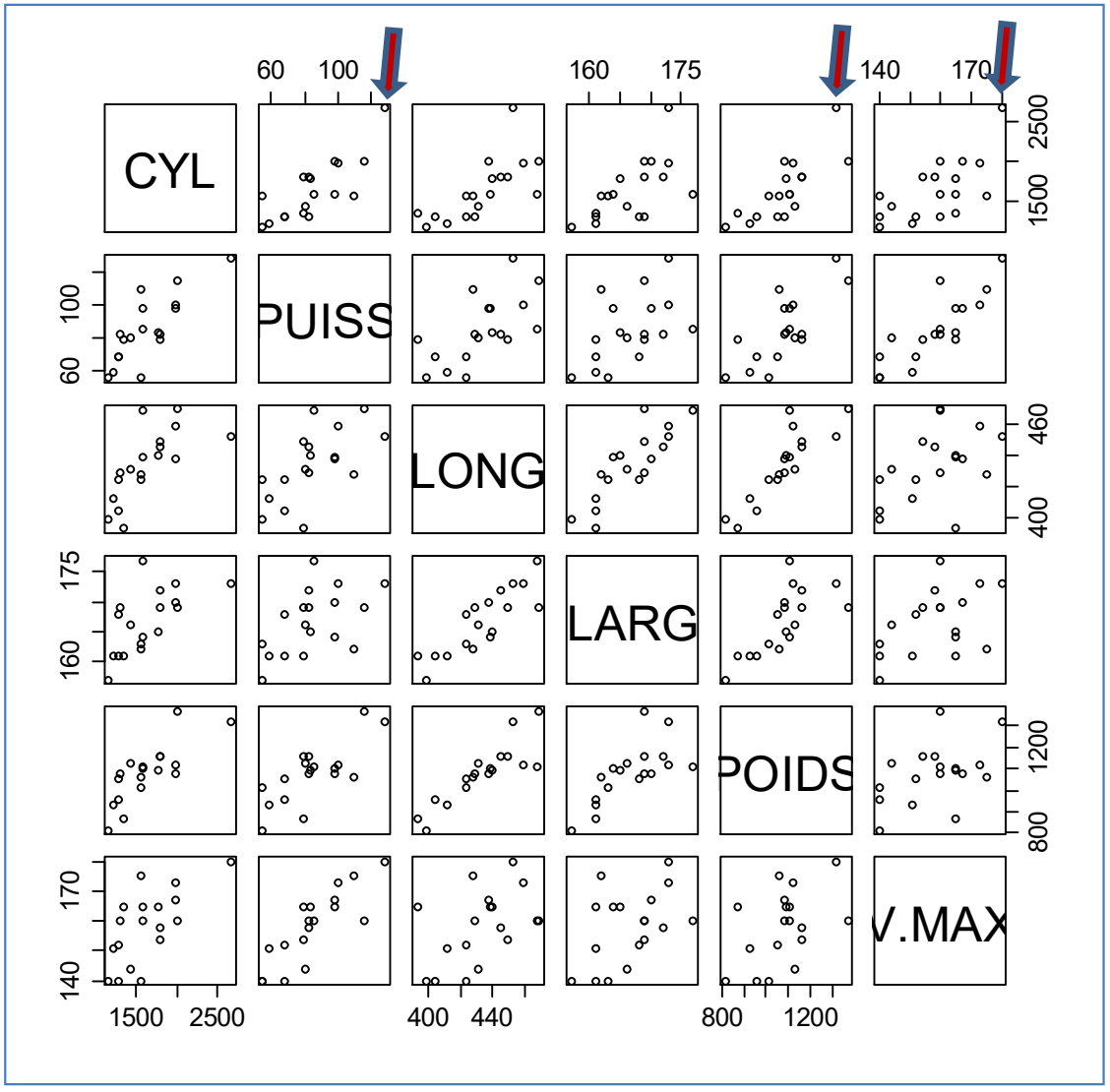
Tableau trié selon CYL

Positionnement des individus (2 variables)



Que faire si on veut prendre en compte (p > 2) variables **simultanément** ?

Positionnement des individus (p >2)



Impossible de créer un nuage à « p » dimensions.

On pourrait croiser les variables 2 à 2, mais :

1. Très difficile de surveiller plusieurs cadrans en même temps.
2. Etiqueter les points rendrait le tout illisible.

Ce type de représentation n'est utile que pour effectuer un diagnostic rapide et repérer les points atypiques.

Ex. Renault 30 : le plus gros moteur, la plus puissante, une des plus lourdes, la plus rapide.

Positionnement des individus – Principe de l'ACP (1) – Notion d'inertie

Principe : Construire un système de représentation de dimension réduite (q << p) qui préserve les **distances** entre les individus. On peut la voir comme une compression avec perte (contrôlée) de l'information.

Distance euclidienne entre 2 individus (i, i')

$$d^2(i,i') = \sum_{j=1}^{p} \left(x_{ij} - x_{i'j}\right)^2$$

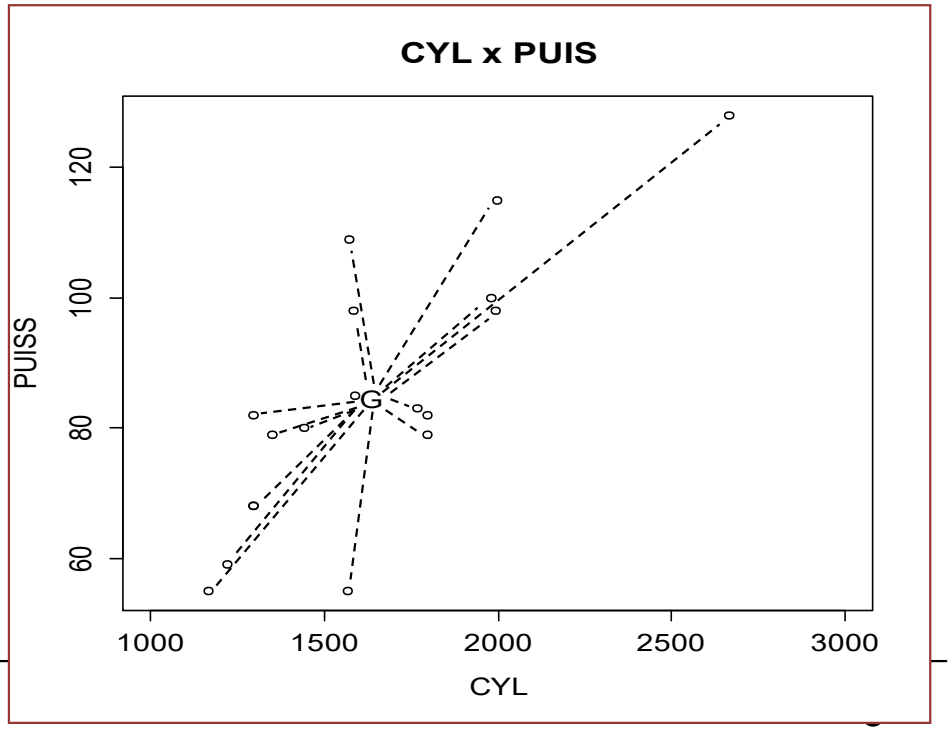
Un critère global : distance entre l'ensemble des individus pris 2 à 2, **inertie** du nuage de points dans l'espace originel. Elle traduit la quantité d'information disponible.

$$I_p = \frac{1}{2n^2} \sum_{i=1}^{n} \sum_{i'=1}^{n} d^2(i,i')$$

Autre écriture de l'inertie : écart par rapport au barycentre G (vecteur constitué des moyennes des p variables)

$$I_p = \frac{1}{n} \sum_{i=1}^{n} d^2(i,G)$$

L'inertie indique la dispersion autour du barycentre, c'est une variance multidimensionnelle (calculée sur p dimensions)



Habituellement on (a) centre et (b) réduit les variables. On parle d'ACP normée.

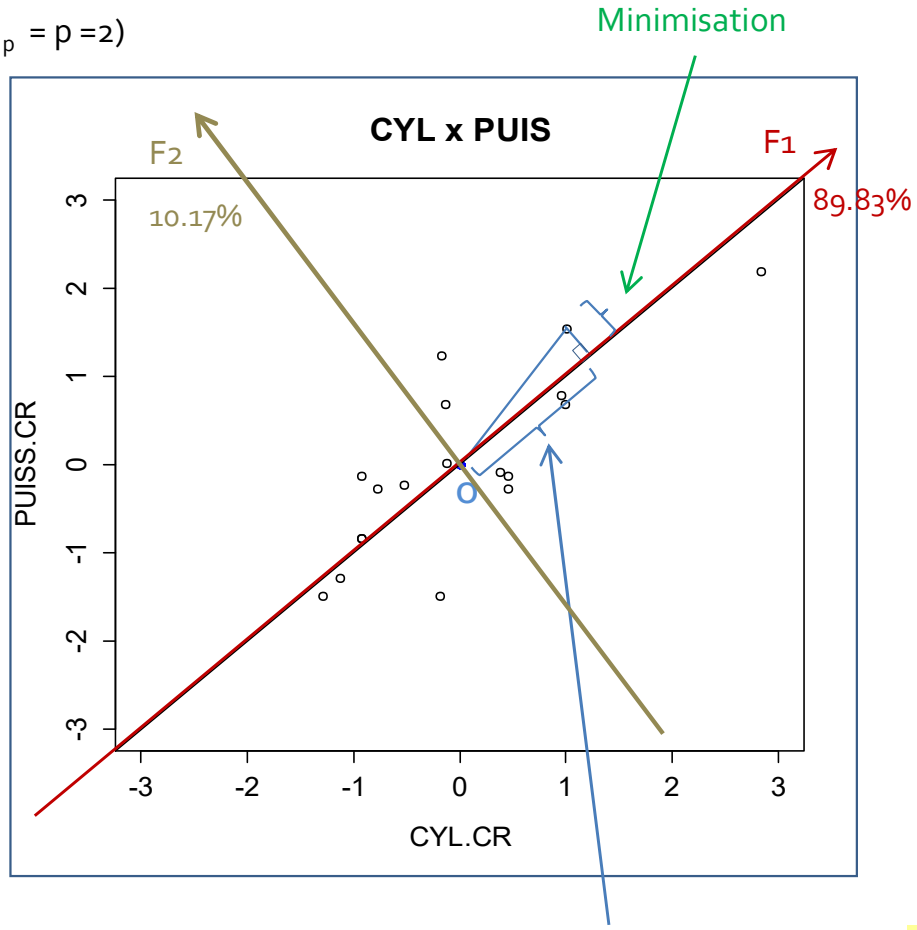
(a) Pour que G soit situé à l'origine [obligatoire]

(b) Pour rendre comparables les variables exprimées sur des échelles (unités) différentes [non obligatoire]

$$z_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j} \quad \left\{ \begin{array}{l} \bar{x}_j = \frac{1}{n}\sum_{i=1}^{n} x_{ij} \\ s_j^2 = \frac{1}{n}\sum_{i=1}^{n}(x_{ij} - \bar{x}_j)^2 \end{array} \right.$$

Cas particulier de 2 variables c.r.

($l_p = p = 2$)



(1) Trouver la première composante F_1 qui maximise l'écartement global des points par rapport à l'origine :

$$\lambda_1 = \frac{1}{n}\sum_{i=1}^{n} F_{i1}^2 = 1.796628$$

$\frac{\lambda_1}{I_p} = 89.83\%$ est la part d'inertie expliquée par le 1er axe factoriel (ou 1ère composante)

(2) Trouver la 2nde composante F_2 qui traite l'inertie non-expliquée (résiduelle) par F_1 (par conséquent, F_2 est non corrélée avec F_1)

$$\lambda_2 = \frac{1}{n}\sum_{i=1}^{n} F_{i2}^2 = 0.203372 \qquad \left(\frac{\lambda_2}{I_p} = 10.17\%\right)$$

(3) Et bien évidemment : $\sum_{k=1}^{p} \lambda_k = 1.797 + 0.203 = 2 = I_p$

Les inerties expliquées s'additionnent. Prendre tous les « p » facteurs possibles permet de récupérer toute l'information disponible !

Préservation des proximités dans le repère réduit

(1) Les proximités entre individus sont préservées si on prend un nombre q de composantes suffisamment représentatives (en terme de % d'inertie exprimée)

(2) Si on prend les « p » facteurs, on retrouve les distances dans le repère originel

Distances dans le repère originel (variables centrées et réduites)

$$d^2(1,2) = (-1.2814 - (-1.1273))^2 + (-1.4953 - (-1.2933))^2 = 0.06455$$
$$d^2(2,6) = 1.14415$$
$$d^2(1,6) = 1.72529$$

	Modele	CYL	PUISS
1	Toyota Corolla	-1.2814	-1.4953
2	Citroen GS Club	-1.1273	-1.2933
3	Simca 1300	-0.9292	-0.8389
4	Lada 1300	-0.9292	-0.8389
5	Lancia Beta	-0.9209	-0.1319
6	Alfasud TI	-0.7751	-0.2834
7	Rancho	-0.5219	-0.2329
8	Renault 16 TL	-0.1835	-1.4953
9	Alfetta 1.66	-0.1697	1.2316
10	Fiat 132	-0.1284	0.6761
11	Audi 100	-0.1202	0.0196
12	Mazda 9295	0.3779	-0.0814
13	Peugeot 504	0.4522	-0.2834
14	Princess 1800	0.4577	-0.1319
15	Opel Rekord	0.9558	0.7771
16	Taunus 2000	0.9943	0.6761
17	Datsun 200L	1.0081	1.5346
18	Renault 30	2.8408	2.1911

Données centrées et réduites

	Modele	F1 (89.83%)	F2 (10.17%)
1	Toyota Corolla	1.9635	0.1513
2	Citroen GS Club	1.7117	0.1174
3	Simca 1300	1.2502	-0.0639
4	Lada 1300	1.2502	-0.0639
5	Lancia Beta	0.7444	-0.5580
6	Alfasud TI	0.7484	-0.3477
7	Rancho	0.5337	-0.2044
8	Renault 16 TL	1.1871	0.9276
9	Alfetta 1.66	-0.7509	-0.9909
10	Fiat 132	-0.3873	-0.5689
11	Audi 100	0.0711	-0.0989
12	Mazda 9295	-0.2097	0.3248
13	Peugeot 504	-0.1194	0.5201
14	Princess 1800	-0.2304	0.4169
15	Opel Rekord	-1.2254	0.1263
16	Taunus 2000	-1.1812	0.2250
17	Datsun 200L	-1.7980	-0.3723
18	Renault 30	-3.5581	0.4594

Coordonnées dans le repère factoriel

Si on ne tient compte que de la 1ère composante ($\lambda_1 = 89.83\%$), les distances sont approximées. On constate néanmoins que les proximités sont assez bien respectées (globalement).

$$d^2_{\{F_1\}}(1,2) = (1.9335 - 1.7117)^2 = 0.06340$$
$$d^2_{\{F_1\}}(2,6) = 0.92783$$
$$d^2_{\{F_1\}}(1,6) = 1.147632$$

Si on tient compte des 2 composantes, on retrouve les distances exactes entre les individus.

$$d^2_{\{F_1,F_2\}}(1,2) = (1.9635 - 1.7117)^2 + (0.1513 - 0.1174)^2 = 0.06455$$
$$d^2_{\{F_1,F_2\}}(2,6) = 1.14415$$
$$d^2_{\{F_1,F_2\}}(1,6) = 1.72529$$

Une des questions clés de l'ACP est de définir le nombre de composantes « q » à retenir pour obtenir une approximation suffisamment satisfaisante !!!

Position du problème (2)

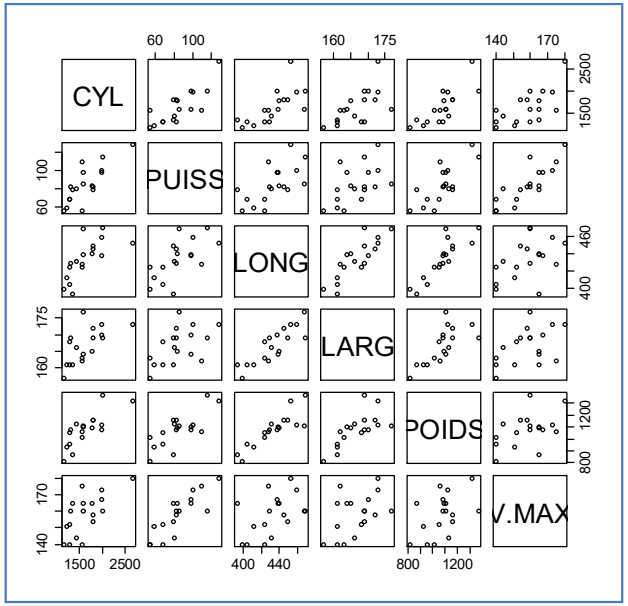
Analyse des relations entre les variables

Le coefficient de corrélation mesure la liaison (linéaire) entre deux variables X_j et X_m

$$r_{jm} = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_{ij} - \bar{x}_j)(x_{im} - \bar{x}_m)}{s_j \times s_m}$$

Matrice des corrélations **R** sur les données « autos »

CORR	CYL	PUISS	LONG	LARG	POIDS	V.MAX
CYL	1	0.797	0.701	0.630	0.789	0.665
PUISS		1	0.641	0.521	0.765	0.844
LONG			1	0.849	0.868	0.476
LARG				1	0.717	0.473
POIDS					1	0.478
V.MAX						1



Elle traduit numériquement ce que l'on peut observer dans les graphiques croisés des variables

On peut essayer de la réorganiser manuellement pour mieux faire apparaître les « blocs » de variables mais….

	POIDS	CYL	PUISS	LONG	LARG	V.MAX
POIDS	1.000	0.789	0.765	0.868	0.717	0.478
CYL		1.000	0.797	0.701	0.630	0.665
PUISS			1.000	0.641	0.521	0.844
LONG				1.000	0.849	0.476
LARG					1.000	0.473
V.MAX						1.000

(1) Ce ne sera jamais parfait

(2) La manipulation est inextricable dès que le nombre de variables est élevé

Construire la première composante F_1 qui permet de maximiser le carré de sa corrélation avec les variables de la base de données

$$\lambda_1 = \sum_{j=1}^{p} r_j^2(F_1)$$

Habituellement, Inertie totale = Somme des variances des variables
Lorsque les données sont réduites (ACP normée), Inertie totale = Trace(R) = p

$$I_p = p$$

⇒ Part d'inertie expliquée par F_1 = $\frac{\lambda_1}{p}$

De nouveau, on observe la décomposition de l'information en composantes non corrélées (orthogonales)

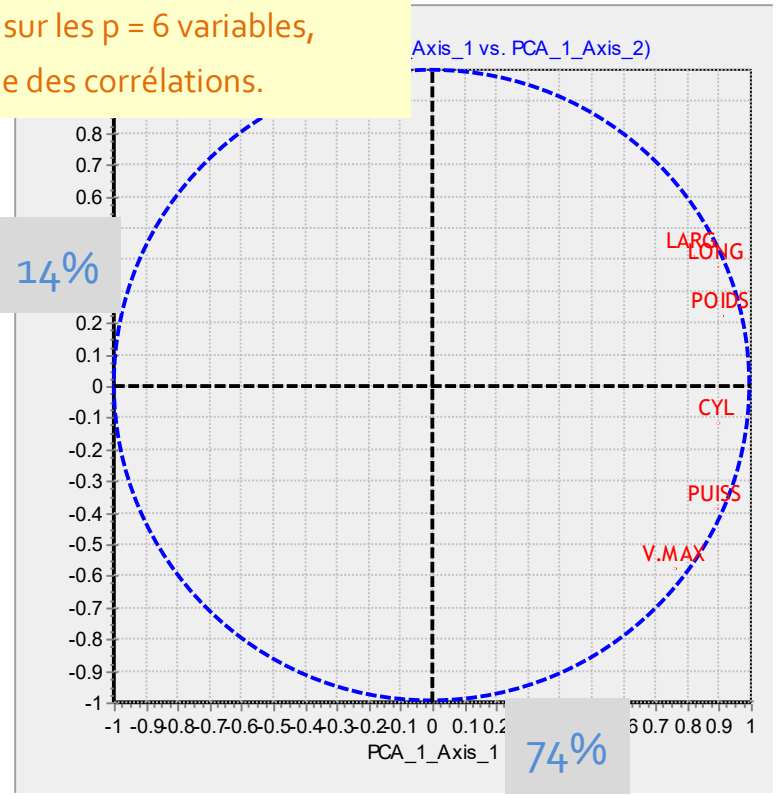
$$\sum_{k=1}^{p} \lambda_k = p$$

Exemple de traitement pour les p = 6 variables de la base de données

Axis	Eigen value	Proportion (%)	Cumulative (%)
1	4.421	73.68%	73.68%
2	0.856	14.27%	87.95%
3	0.373	6.22%	94.17%
4	0.214	3.57%	97.73%
5	0.093	1.55%	99.28%
6	0.043	0.72%	100.00%
Tot.	6	-	-

Relations entres variables – Principe de l'ACP (2) – Approximation des corrélations

ACP sur les p = 6 variables, cercle des corrélations.



Liaison de la variable « poids » avec le 1er axe

$$r_{poids}(F_1) = 0.905 \quad \text{et} \quad r^2_{poids}(F_1) = 0.819$$

La représentation de la variable n'est pas complète, on a besoin d'un second facteur F_2

$$r_{poids}(F_2) = 0.225 \quad \text{et} \quad r^2_{poids}(F_2) = 0.050$$

Si on exploite tous les « p » facteurs

$$\sum_{k=1}^{p} r^2_{poids}(F_k) = 0.819 + 0.050 + \cdots = 1$$

L'ACP produit aussi une approximation dans l'espace des variables (**approximation des corrélations**)

[Ex. si on ne prend en compte que « q = 1 » facteur]

$$\begin{cases} r_{poids,cyl} = 0.789 \\ r_{poids,cyl}(F_1) = \sum_{k=1}^{q=1} r_{poids}(F_k) \times r_{cyl}(F_k) = 0.90519 \times 0.89346 = 0.809 \end{cases}$$

Approximation assez bonne parce que POIDS et CYL sont bien représentées sur le 1er facteur

$$\begin{cases} r_{poids,v.\max} = 0.478 \\ r_{poids,v.\max}(F_1) = 0.90519 \times 0.75471 = 0.683 \end{cases}$$

Approximation mauvaise parce que V.MAX est mal représentée sur le premier facteur [(0.75471^2)=57% de l'information seulement]

Calculs

Les mains dans le cambouis : comment sont obtenus les résultats de l'ACP ?

Objectif des calculs

Construire un ensemble de composantes (F_1, F_2, …, F_k,…), combinaisons linéaires des variables originelles (centrées et réduites), dont on peut apprécier la qualité de restitution de l'information à travers l'inertie reproduite (λ_k)

$$\begin{cases} F_1 = a_{11}z_1 + a_{21}z_2 + \cdots + a_{p1}z_p \ (\lambda_1) \\ \vdots \\ F_k = a_{1k}z_1 + a_{2k}z_2 + \cdots + a_{pk}z_p \ (\lambda_k) \\ \vdots \\ F_p = a_{1p}z_1 + a_{2p}z_2 + \cdots + a_{pp}z_p \ (\lambda_p) \end{cases}$$

Comment obtenir les coefficients « a_{jk} » à partir des données ?

Qui permettent de calculer les coordonnées des individus dans le repère factoriel, et de juger de leur proximité dans les différents plans factoriels

$$F_{ik} = a_{1k}z_{i1} + a_{2k}z_{i2} + \cdots + a_{pk}z_{ip}$$

Valeur de la variable Z_2 (X_2 après centrage et réduction) pour l'individu n°i

Que l'on interprétera en calculant leur corrélations (et autres indicateurs dérivés : CTR et COS²) avec les variables originelles (X_1, X_2, …, X_p)

$$r_{x_j}(F_k)$$

Plus la corrélation est élevée en valeur absolue, plus forte est l'influence de la variable sur le facteur

Calcul via la diagonalisation de la matrice des corrélations

Calcul uniquement dans l'espace des variables,

mais résultats disponibles pour les deux points de vue (individus et variables)

```
#chargement du fichier de données
autos <- read.table(file="autos.txt",sep="\t",row.names=1,header=

#calcul de la matrice des corrélations
autos.cor <- cor(autos)
print(autos.cor)

#trace de la matrice = inertie totale
print(sum(diag(autos.cor)))

#diagonalisation avec la fonction eigen
autos.eigen <- eigen(autos.cor)
print(autos.eigen)

#calcul des corrélations des variables avec les composantes
cor.factors <- NULL
for (j in 1:ncol(autos)){
  rf <- sqrt(autos.eigen$values[j])*autos.eigen$vectors[,j]
  cor.factors <- cbind(cor.factors,rf)
}
rownames(cor.factors) <- colnames(autos)
colnames(cor.factors) <- paste("F",1:ncol(autos),sep="")

#affichage des 2 premières composantes seulement
print(cor.factors[,1:2])
```

	CYL	PUISS	LONG	LARG	POIDS	V.MAX
CYL	1.0000000	0.7966277	0.7014619	0.6297572	0.7889520	0.6649340
PUISS	0.7966277	1.0000000	0.6413624	0.5208320	0.7652930	0.8443795
LONG	0.7014619	0.6413624	1.0000000	0.8492664	0.8680903	0.4759285
LARG	0.6297572	0.5208320	0.8492664	1.0000000	0.7168739	0.4729453
POIDS	0.7889520	0.7652930	0.8680903	0.7168739	1.0000000	0.4775956
V.MAX	0.6649340	0.8443795	0.4759285	0.4729453	0.4775956	1.0000000

```
> print(sum(diag(autos.cor)))
[1] 6
```

Valeurs propres = λ_k

```
> print(autos.eigen)
$values
[1] 4.42085806 0.85606229 0.37306608 0.21392209 0.09280121 0.04329027

$vectors
           [,1]       [,2]        [,3]       [,4]       [,5]        [,6]
[1,] -0.4249360  0.1241911 -0.35361252  0.8077865 -0.1515800 -0.05889517
[2,] -0.4217944  0.4157739 -0.18492049 -0.3577920  0.2937346 -0.63303302
[3,] -0.4214599 -0.4118177  0.06763394 -0.2797523 -0.7305690 -0.19029153
[4,] -0.3869222 -0.4460870  0.60486812  0.2115694  0.4781901 -0.10956624
[5,] -0.4305120 -0.2426758 -0.48439601 -0.3017114  0.3045584  0.58081220
[6,] -0.3589443  0.6198626  0.48547226 -0.0735743 -0.1886551  0.45852167
```

Vecteurs propres = a_{jk}

	F1	F2
CYL	-0.8934635	0.1149061
PUISS	-0.8868580	0.3846891
LONG	-0.8861548	-0.3810287
LARG	-0.8135364	-0.4127359
POIDS	-0.9051875	-0.2245325
V.MAX	-0.7547104	0.5735194

Corrélations variables x facteurs

$$r_{x_j}(F_k) = \sqrt{\lambda_k} \times a_{jk}$$

Calcul via la décomposition en valeurs singulières de la matrice des données centrées et réduites

Montre bien le caractère dual de l'analyse

z_{ij}

```
> print(head(autos.cr,3))
                CYL       PUISS       LONG       LARG      POIDS      V.MAX
Alfasud TI -0.7750989 -0.28335818 -1.8850808 -1.0973453 -1.5690068  0.5697604
Audi 100   -0.1201633  0.01963869  1.6058095  2.0010414  0.2341614  0.1459717
Simca 1300 -0.9292014 -0.83885242 -0.4421794  0.2581989 -0.2166306 -0.5320903
```

```
#affichage des 3 premières obs. de Z
print(head(autos.cr,3))

#décomposition en valeurs singulières
svd.autos <- svd(autos.cr)
print(svd.autos,digits=3)

#calcul des inerties associées aux composantes
print(svd.autos$d^2/nrow(autos))
```

On obtient les coordonnées factorielles des individus avec $F_{ik} = \delta_k \times u_{ik}$

```
> print(svd.autos,digits=3)
$d
[1] 8.921 3.925 2.591 1.962 1.292 0.883

$u
         [,1]    [,2]     [,3]     [,4]    [,5]    [,6]
 [1,] -0.2398 -0.4549 -0.22068 -0.10290  0.2332 -0.0611
 [2,]  0.1750  0.3890 -0.50756  0.10771 -0.1149  0.3707
 [3,] -0.1255  0.1718 -0.17620  0.08542  0.2904 -0.3079
 [4,] -0.2885 -0.0288 -0.05733  0.00884 -0.1755 -0.2985
 [5,]  0.0480 -0.1772  0.07459  0.31991 -0.2039  0.0421
 [6,] -0.0341  0.0500 -0.26079  0.28331  0.3444 -0.2267
 [7,]  0.0767  0.2377  0.09911 -0.10352 -0.1614 -0.1743
 [8,] -0.2184  0.2498  0.23909 -0.32122 -0.2268 -0.1231
 [9,]  0.4943 -0.2710  0.22904 -0.43176  0.2901 -0.0498
[10,] -0.4468 -0.0602  0.11698 -0.13511 -0.2154  0.3726
[11,]  0.0491 -0.4872 -0.00963  0.38675 -0.1301  0.0614
[12,]  0.1141  0.2144 -0.08359 -0.15463  0.1430 -0.2095
[13,]  0.3297  0.1424  0.48005  0.39350 -0.0421  0.0649
[14,]  0.1474 -0.1239 -0.10906 -0.29671  0.0516  0.2867
[15,] -0.0775  0.2287  0.24250  0.18231  0.2918  0.1377
[16,]  0.0432 -0.0907  0.02917 -0.05244 -0.4078 -0.3838
[17,]  0.2567 -0.0266 -0.30732 -0.12044 -0.2619  0.1775
[18,] -0.3036  0.0366  0.22163 -0.04902  0.2954  0.3211

$v
      [,1]   [,2]    [,3]    [,4]   [,5]    [,6]
[1,] 0.425 -0.124  0.3536 -0.8078 -0.152  0.0589
[2,] 0.422 -0.416  0.1849  0.3578  0.294  0.6330
[3,] 0.421  0.412 -0.0676  0.2798 -0.731  0.1903
[4,] 0.387  0.446 -0.6049 -0.2116  0.478  0.1096
[5,] 0.431  0.243  0.4844  0.3017  0.305 -0.5808
[6,] 0.359 -0.620 -0.4855  0.0736 -0.189 -0.4585
```

Principe de la SVD

$$Z = U \Delta V^T \quad \text{avec} \quad \begin{cases} Z\vec{v}_k = \delta_k \vec{u}_k \\ Z^T\vec{u}_k = \delta_k \vec{v}_k \end{cases}$$

V correspond aux vecteurs propres c.-à-d. les coef. a_{jk}

Calcul des inerties : $\lambda_k = \dfrac{\delta_k^2}{n}$

```
> print(svd.autos$d^2/nrow(autos))
[1] 4.42085806 0.85606229 0.37306608 0.21392209 0.09280121 0.04329027
```

Pratique de l'ACP

Que lire et comment lire les résultats de l'ACP ?

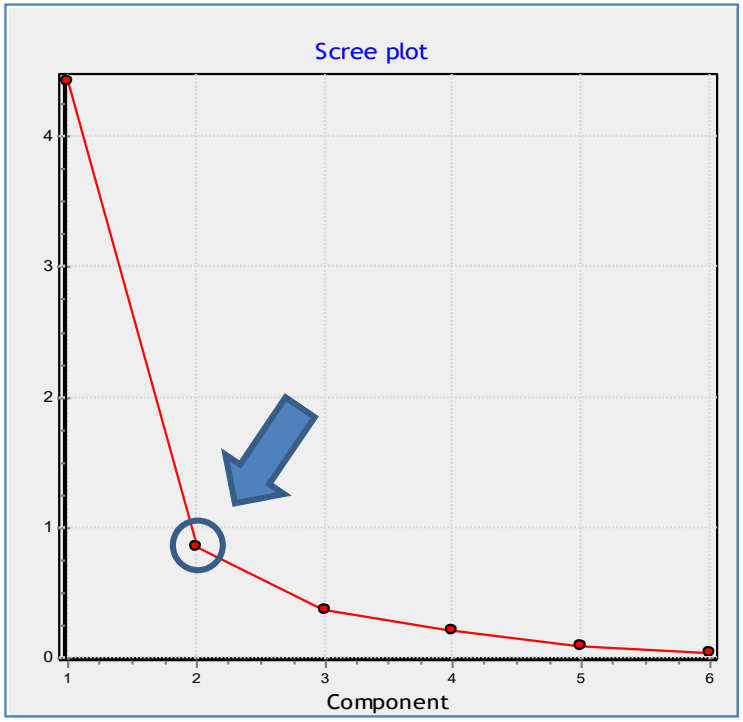
Détermination du nombre de composantes à retenir

(1) Tableau des valeurs propres

Axis	Eigen value	Difference	Proportion (%)	Cumulative (%)
1	4.420858	3.564796	73.68%	73.68%
2	0.856062	0.482996	14.27%	87.95%
3	0.373066	0.159144	6.22%	94.17%
4	0.213922	0.121121	3.57%	97.73%
5	0.092801	0.049511	1.55%	99.28%
6	0.04329	-	0.72%	100.00%
Tot.	6	-	-	-

Indications : (1) sur l'importance des composantes, (2) sur l'évolution de l'importance cumulée, (3) sur la qualité de l'information restituée par les « q » premiers facteurs.

(2) Eboulis des valeurs propres : « scree plot »



« Règle du coude » de Cattell, négliger les composantes qui emmènent peu d'informations additionnelles. Très performante lorsqu'il y a des « blocs » de variables. Fournit surtout des scénarios de solutions.

Problème : Intégrer le coude dans la sélection ? Ici, q = 2 ou q = 1 ? Tout dépend de la valeur propre associée au coude, si elle est faible, il faut exclure la composante associée.

Mais, en pratique, (a) il faut au moins « q = 2 » afin de pouvoir réaliser les représentations graphiques; (b) il faut aussi pouvoir interpréter les composantes.

Règle de Kaiser-Guttman : si les variables sont indépendantes deux à deux, les valeurs propres λ_k seraient toutes égales à 1.

Remarque 1 : cette règle ne tient pas compte du tout des caractéristiques des données.

Remarque 2 : On peut aussi voir le seuil « 1 » comme la moyenne des valeurs propres.

Axis	Eigen value	Difference	Proportion (%)	Cumulative (%)
1	**4.420858**	3.564796	73.68%	73.68%
2	0.856062	0.482996	14.27%	87.95%
3	0.373066	0.159144	6.22%	94.17%
4	0.213922	0.121121	3.57%	97.73%
5	0.092801	0.049511	1.55%	99.28%
6	0.04329	-	0.72%	100.00%
Tot.	6	-	-	-

Règle de Karlis-Saporta-Spinaki : rendre la règle plus stricte en tenant compte des caractéristiques (n et p) des données.

$$\lambda > 1 + 2\sqrt{\frac{p-1}{n-1}}$$

A droite, nous avons 2 x écart-type des v.p. sous Ho ≈ un test unilatéral à 5%

$$seuil = 1 + 2\sqrt{\frac{p-1}{n-1}} = 1 + 2\sqrt{\frac{6-1}{18-1}} = 2.08465$$

Axis	Eigen value	Difference	Proportion (%)	Cumulative (%)
1	**4.420858**	3.564796	73.68%	73.68%
2	0.856062	0.482996	14.27%	87.95%
3	0.373066	0.159144	6.22%	94.17%
4	0.213922	0.121121	3.57%	97.73%
5	0.092801	0.049511	1.55%	99.28%
6	0.04329	-	0.72%	100.00%
Tot.	6	-	-	-

Test des « bâtons brisés » de Frontier (1976) et Legendre-Legendre (1983) : si l'inertie était répartie aléatoirement sur les axes, la distribution des v.p. suivrait la loi des « bâtons brisés ».

Problème : les tables sont rarement accessibles. Heureusement les valeurs critiques à 5% peuvent être obtenues très facilement.

$$b_k = \sum_{m=k}^{p} \frac{1}{m}$$

La composante est validée si : $\lambda_k > b_k$

$$b_1 = \frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5} + \frac{1}{6} = 2.45$$

Eigenvalues - Significance		
Axis	Eigenvalue	Broken-stick critical values
1	4.420858	2.45
2	0.856062	1.45
3	0.373066	0.95
4	0.213922	0.616667
5	0.092801	0.366667
6	0.04329	0.166667

$$b_3 = \frac{1}{3} + \frac{1}{4} + \frac{1}{5} + \frac{1}{6} = 0.95$$

Toutes les approches sont cohérentes : q = 1 seul facteur semble suffire dans cette étude. Par commodité (hum, pas si sûr, cf. interprétation et rotation des axes), on en choisira q = 2.

Caractérisation des composantes par les variables

Analyse des relations entre les variables via les composantes

Corrélation : degré de liaison de la variable avec la composante

$$\sum_{j=1}^{p} r_{x_j}^2(F_k) = \lambda_k$$

Contributions : influence de la variable dans la définition de la composante (rarement fournie car redondante avec CORR et COS²)

$$CTR_{jk} = \frac{r_{x_j}^2(F_k)}{\lambda_k}; \sum_{j=1}^{p} CTR_{jk} = 1$$

Cosinus carré : qualité de représentation de la variable sur la composante. On peut la cumuler sur les q premières composantes.

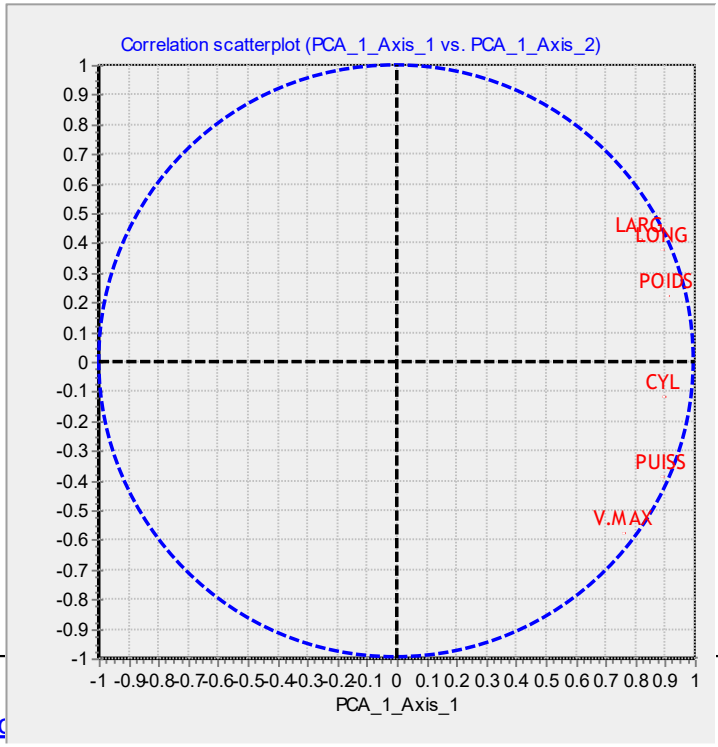
$$COS_{jk}^2 = r_{x_j}^2(F_k)$$

$$COS_{jq}^2 = \sum_{k=1}^{q} COS_{jk}^2$$

$$\sum_{k=1}^{p} COS_{jk}^2 = 1$$

	Axis_1			Axis_2		
	Corr.	CTR (%)	COS² (%)	Corr.	CTR (%)	COS² (%)
POIDS	0.905	19%	*82 % (82 %)*	0.225	6%	*5 % (87 %)*
CYL	0.893	18%	*80 % (80 %)*	-0.115	2%	*1 % (81 %)*
PUISS	0.887	18%	*79 % (79 %)*	-0.385	17%	*15 % (93 %)*
LONG	0.886	18%	*79 % (79 %)*	0.381	17%	*15 % (93 %)*
LARG	0.814	15%	*66 % (66 %)*	0.413	20%	*17 % (83 %)*
V.MAX	0.755	13%	*57 % (57 %)*	-0.574	38%	*33 % (90 %)*
Var. Expl.	4.42086		*74 % (74 %)*	0.85606		*14 % (88 %)*

On utilise souvent le « cercle des corrélations » pour obtenir une vision synthétique immédiate.



On observe (axe 1 : 74%) un « effet taille » marqué, que l'on peut lier à « l'encombrement / gamme » des véhicules ; mais aussi (axe 2 : 14%), une caractérisation par les performances (sportivité).

A nombre de composantes fixé, on peut comparer les corrélations brutes calculées (en bleu) sur les données originelles, et celles estimées à partir du repère factoriel (en vert). Nous avons choisi q = 2 pour les données « AUTOS ».

$$\hat{r}_q(x_j, x_{j'}) = \sum_{k=1}^{q} r_{x_j}(F_k) \times r_{x_{j'}}(F_k)$$

Entre parenthèses la différence entre les corrélations ➔

Original, reproduced and residual correlations

	CYL	PUISS	LONG	LARG	POIDS	V.MAX
CYL	-	0.7966 0.8366 (-0.0400)	0.7015 0.7480 (-0.0465)	0.6298 0.6794 (-0.0497)	0.7890 0.7830 (0.0060)	0.6649 0.7402 (-0.0753)
PUISS	0.7966 0.8366 (-0.0400)	-	0.6414 0.6393 (0.0020)	0.5208 0.5627 (-0.0419)	0.7653 0.7164 (0.0489)	0.8444 0.8899 (-0.0456)
LONG	0.7015 0.7480 (-0.0465)	0.6414 0.6393 (0.0020)	-	0.8493 0.8782 (-0.0289)	0.8681 0.8877 (-0.0196)	0.4759 0.4503 (0.0257)
LARG	0.6298 0.6794 (-0.0497)	0.5208 0.5627 (-0.0419)	0.8493 0.8782 (-0.0289)	-	0.7169 0.8291 (-0.1122)	0.4729 0.3773 (0.0957)
POIDS	0.7890 0.7830 (0.0060)	0.7653 0.7164 (0.0489)	0.8681 0.8877 (-0.0196)	0.7169 0.8291 (-0.1122)	-	0.4776 0.5544 (-0.0768)
V.MAX	0.6649 0.7402 (-0.0753)	0.8444 0.8899 (-0.0456)	0.4759 0.4503 (0.0257)	0.4729 0.3773 (0.0957)	0.4776 0.5544 (-0.0768)	-

Attribute	Axis_1		Axis_2	
-	Corr.	% (Tot. %)	Corr.	% (Tot. %)
CYL	0.89346	80 % (80 %)	-0.11491	1 % (81 %)
PUISS	0.88686	79 % (79 %)	-0.38469	15 % (93 %)
LONG	0.88615	79 % (79 %)	0.38103	15 % (93 %)
LARG	0.81354	66 % (66 %)	0.41274	17 % (83 %)
POIDS	0.90519	82 % (82 %)	0.22453	5 % (87 %)
V.MAX	0.75471	57 % (57 %)	-0.57352	33 % (90 %)
Var. Expl.	4.42086	74 % (74 %)	0.85606	14 % (88 %)

L'approximation sera d'autant meilleure que les variables sont bien représentées dans le repère sélectionné.

COS² des variables cumulé pour les 2 premières composantes

Caractérisation des composantes par les individus

Analyse des proximités entre individus via leurs coordonnées factorielles

N.B. $I_p = \frac{1}{n}\sum_{i=1}^{n} d^2(i,G) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{p} z_{ij}^2 = \frac{1}{n}\sum_{i=1}^{n} d_i^2$

d_i^2 indique la part de l'individu dans l'inertie totale (dans l'espace originel – variables c.r.). C'est le carré de la distance à l'origine.

Caractérisation des facteurs à l'aide des individus – Coordonnées, contributions et cos²

(1) Coordonnée factorielle de l'individu F_{ik} (permet de situer le positionnement relatif des observations).

(4) Les cos² s' additionnent. Qualité des représentations sur les q = 2 premiers facteurs.

Lecture : Les véhicules se caractérisent par l'encombrement (axe 1, illustrés par les véhicules {9, 10, 13}) et la performance (axe 2, avec surtout {1, 2, 11}).

Remarque : {6, 16 et 5} sont mal représentés sur les q = 2 premières composantes parce qu'ils ne se distinguent ni par l' encombrement (proche de la moyenne) ni par la performance (se situent dans la moyenne).

		Axe 1			Axe 2			
N°	Modele	Coord.	CTR	Cos²	Coord.	CTR	Cos²	SUM(COS²)
1	Alfasud TI	-2.139	6%	56%	-1.786	21%	39%	94%
2	Audi 100	1.561	3%	37%	1.527	15%	35%	71%
3	Simca 1300	-1.119	2%	58%	0.675	3%	21%	79%
4	Citroen GS Club	-2.574	8%	98%	-0.113	0%	0%	98%
5	Fiat 132	0.428	0%	16%	-0.696	3%	41%	57%
6	Lancia Beta	-0.304	0%	8%	0.196	0%	3%	12%
7	Peugeot 504	0.684	1%	31%	0.933	6%	58%	88%
8	Renault 16 TL	-1.948	5%	67%	0.980	6%	17%	84%
9	Renault 30	4.410	24%	89%	-1.064	7%	5%	94%
10	Toyota Corolla	-3.986	20%	98%	-0.236	0%	0%	98%
11	Alfetta 1.66	0.438	0%	4%	-1.912	24%	82%	86%
12	Princess 1800	1.018	1%	53%	0.842	5%	36%	89%
13	Datsun 200L	2.941	11%	78%	0.559	2%	3%	81%
14	Taunus 2000	1.315	2%	70%	-0.487	2%	10%	80%
15	Rancho	-0.691	1%	24%	0.898	5%	41%	65%
16	Mazda 9295	0.386	0%	22%	-0.356	1%	19%	40%
17	Opel Rekord	2.290	7%	86%	-0.104	0%	0%	86%
18	Lada 1300	-2.709	9%	93%	0.144	0%	0%	93%

(2) Contribution : indique l'influence de l'individu dans la définition du facteur

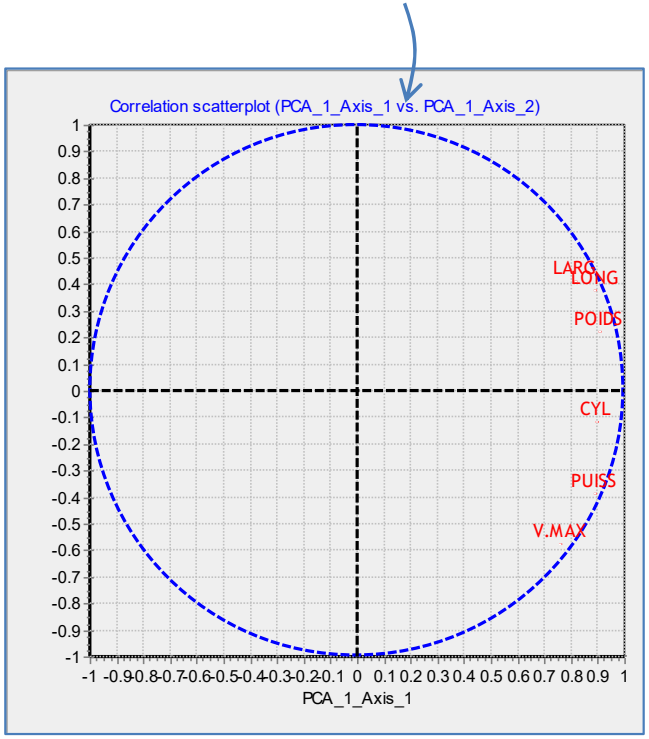
$$CTR_{ik} = \frac{F_{ik}^2}{n \times \lambda_k}; \sum_{i=1}^{n} CTR_{ik} = 1$$

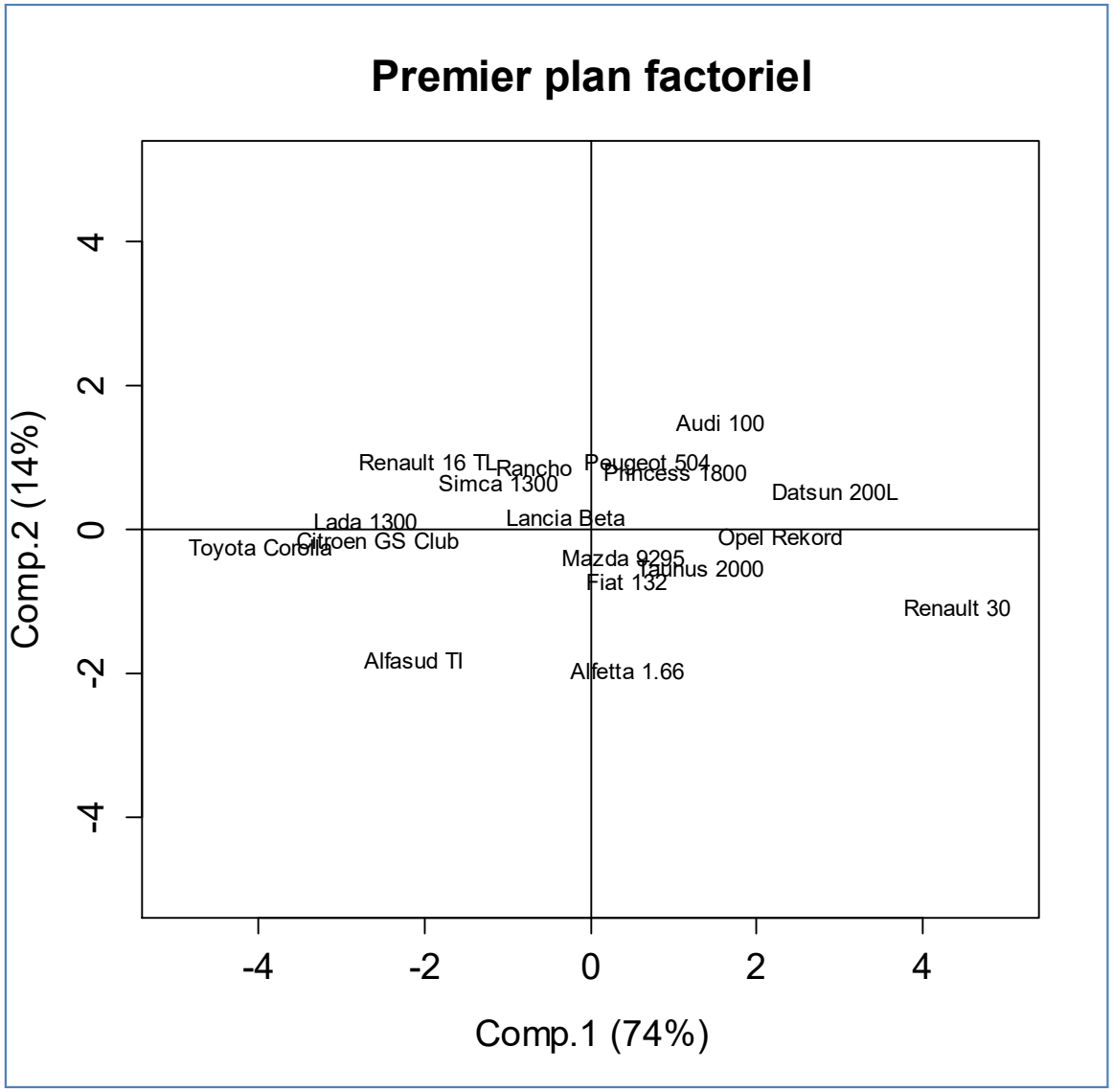
(3) Cos² : indique la qualité de la représentation de l'individu sur le facteur (fraction de son inertie restituée par le facteur)

$$COS_{ik}^2 = \frac{F_{ik}^2}{d_i^2}; \sum_{j=1}^{p} COS_{ik}^2 = 1$$

Ce graphique fait en très grande partie la popularité de l'ACP. On peut y juger visuellement des proximités (dissemblances) entre les individus.

Et on peut comprendre le pourquoi des proximités en considérant dans le même temps le cercle des corrélations.





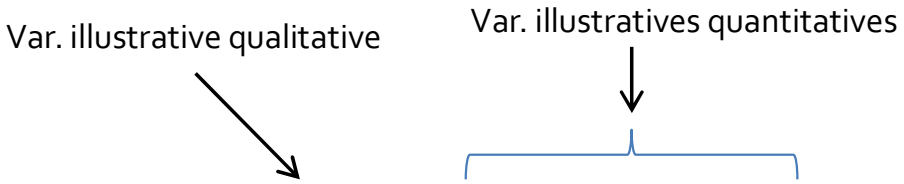
Remarque : certains proposent de mêler les deux représentations dans un graphique dit « biplot ». Attention, les proximités individus-variables n'ont pas vraiment de sens. Ce sont les directions qui importent dans ce cas.

Variables illustratives

Renforcer l'interprétation des composantes

Variables non utilisées pour la construction des composantes. Mais utilisées après coup pour mieux comprendre / commenter les résultats.

Ex. Les caractéristiques intrinsèques des véhicules sont les variables actives (largeur, poids, puissance, etc.). En illustratives, on utilise des variables introduisant des considérations subjectives (prix, gamme) ou calculées après coup pour une meilleure interprétation (rapport poids/puissance).

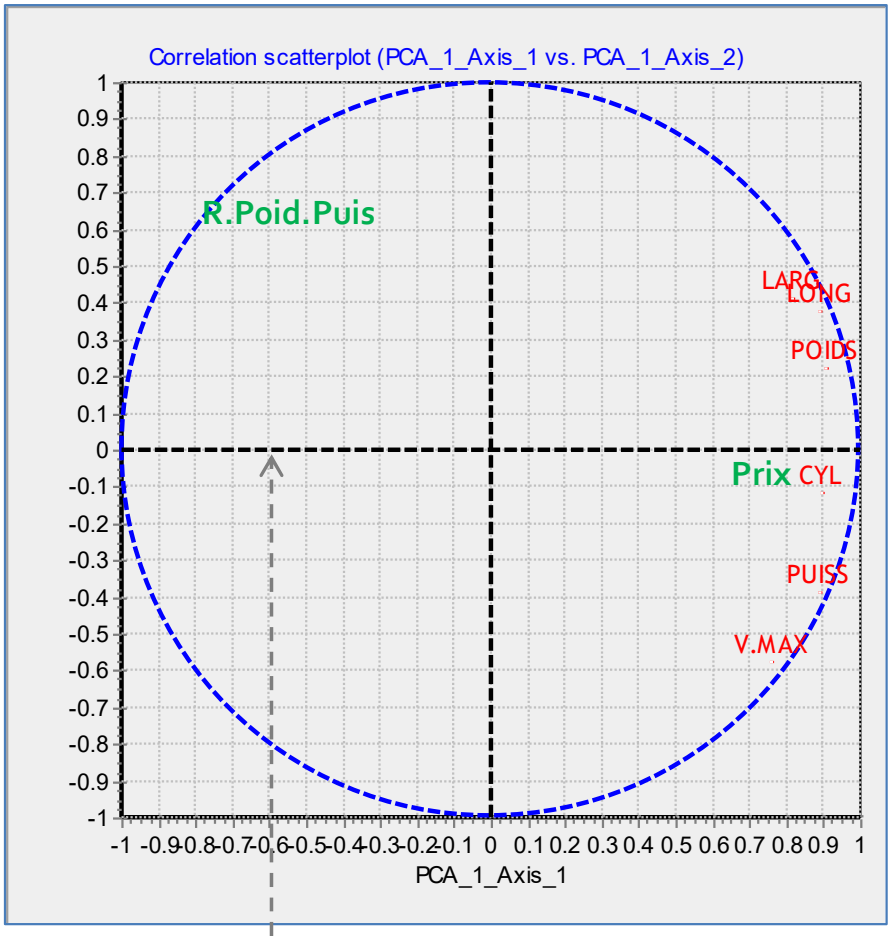


Modele	FINITION	PRIX	R.POID.PUIS
Alfasud TI	2_B	30570	11.01
Audi 100	3_TB	39990	13.06
Simca 1300	1_M	29600	15.44
Citroen GS Club	1_M	28250	15.76
Fiat 132	2_B	34900	11.28
Lancia Beta	3_TB	35480	13.17
Peugeot 504	2_B	32300	14.68
Renault 16 TL	2_B	32000	18.36
Renault 30	3_TB	47700	10.31
Toyota Corolla	1_M	26540	14.82
Alfetta-1.66	3_TB	42395	9.72
Princess-1800	2_B	33990	14.15
Datsun-200L	3_TB	43980	11.91
Taunus-2000	2_B	35010	11.02
Rancho	3_TB	39450	14.11
Mazda-9295	1_M	27900	13.19
Opel-Rekord	2_B	32700	11.20
Lada-1300	1_M	22100	14.04

Variables illustratives quantitatives

$$r_y(F_k) = \frac{\frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})(F_{ik} - \bar{F}_k)}{s_y \times s_{F_k}} = \frac{\frac{1}{n}\sum_{i=1}^{n} F_{ik}(y_i - \bar{y})}{s_y \times \sqrt{\lambda_k}}$$

Calculer les corrélations des variables supplémentaires avec les facteurs. c.-à-d. calculer le coefficient de corrélation entre les coordonnées des « n » individus sur les facteurs et les valeurs prises par la variable illustrative. Il est possible de les placer dans le cercle des corrélations.



CORR	Comp.1	Comp.2
PRIX	0.772	-0.087
R.POID.PUIS	-0.589	0.673

Tester la « significativité » du lien avec la statistique basée sur la transformation de Fisher

$$u_y = \sqrt{n-3} \times \left(\frac{1}{2} \ln \frac{1+r}{1-r} \right)$$

Lien significatif à (~) 5% si

$$|u_y| \geq 2$$

SIGNIF.	Comp.1	Comp.2
PRIX	**3.975**	-0.337
R.POID.PUIS	**-2.619**	**3.158**

Le rapport poids/puissance n'est pas lié positivement avec le poids parce que les voitures lourdes sont comparativement plus puissantes.

$$\mu_{gk} = \frac{1}{n_g} \sum_{i:y_i=g} F_{ik}$$

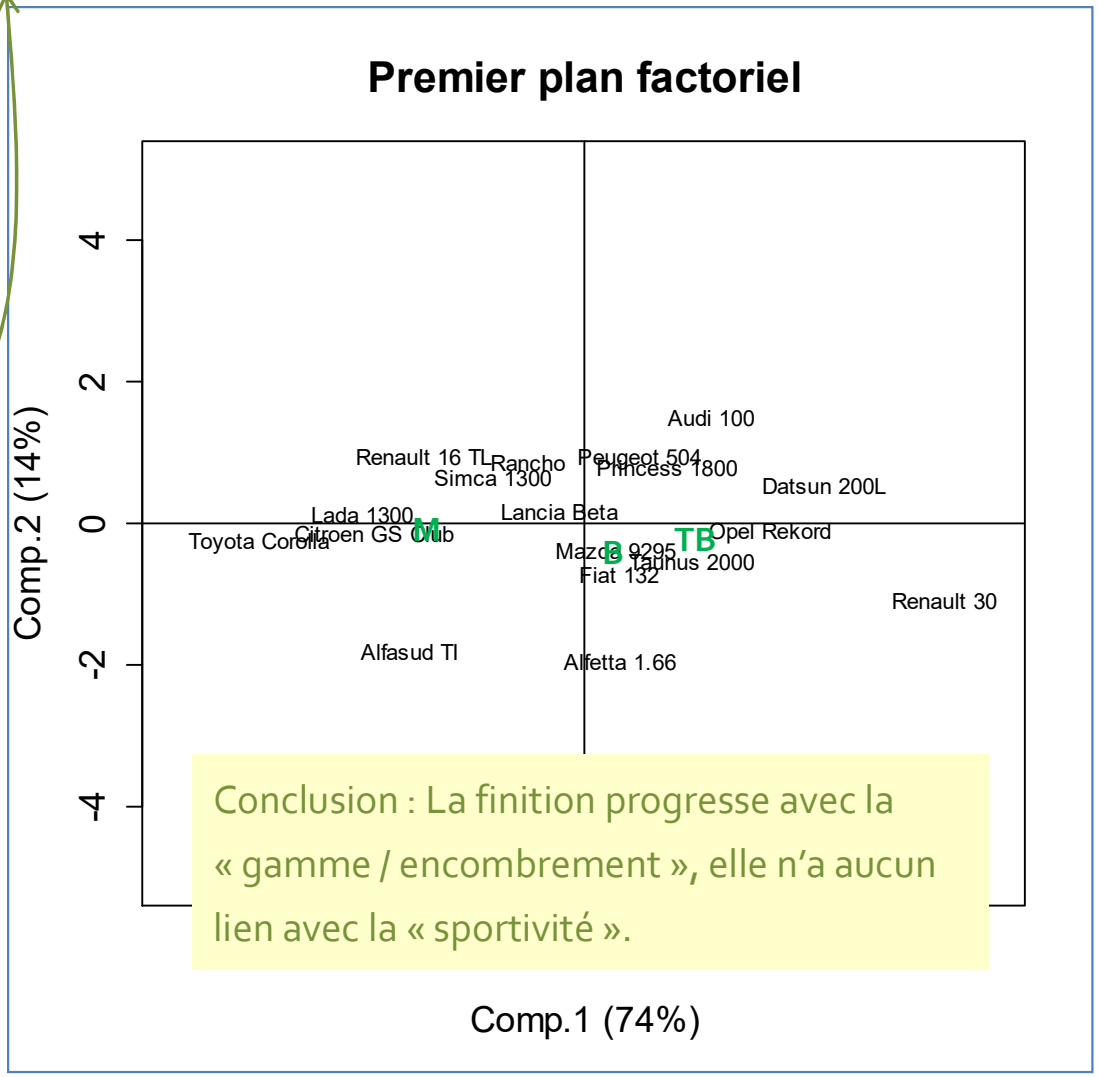
		Comp.1		Comp.2	
FINITION	n_g	**Moyenne**	**Valeur.Test**	**Moyenne**	**Valeur.Test**
1_M	5	-2.0004	-2.43	0.0226	0.06
2_B	7	0.2353	0.37	-0.0453	-0.16
3_TB	6	1.3924	1.93	0.0340	0.11

Comparer les moyennes des composantes conditionnellement aux groupes définis par les modalités de la variable illustrative qualitative.

Possibilité de tester la significativité de l'écart par rapport à l'origine (moyenne des composantes = 0) avec la « valeur test » (Morineau, 1984).

$$VT_{gk} = \frac{\mu_{gk} - \bar{F}_k}{\sqrt{\frac{n-n_g}{n-1} \times \frac{s^2_{F_k}}{n_g}}} = \frac{\mu_{gk} - 0}{\sqrt{\frac{n-n_g}{n-1} \times \frac{\lambda_k}{n_g}}}$$

Ecart significatif à (~) 5% si $|VT_{gk}| \geq 2$



Remarque : On pourrait également s'appuyer sur l'ANOVA pour comparer les moyennes, et/ou calculer le rapport de corrélation.

Individus illustratifs (supplémentaires)

Positionner de nouveaux individus

Plusieurs raisons possibles :

1. Des individus collectés après coup que l'on aimerait situer par rapport à ceux de l'échantillon d'apprentissage (les individus actifs).
2. Des individus appartenant à une population différente (ou spécifique) que l'on souhaite positionner.
3. Des observations s'avérant atypiques ou trop influentes dans l'ACP que l'on a préféré écarter. On veut maintenant pouvoir juger de leur positionnement par rapport aux individus actifs.

Modele	CYL	PUISS	LONG	LARG	POIDS	V.MAX
Peugeot 604	2664	136	472	177	1410	180
Peugeot 304 S	1288	74	414	157	915	160

Plutôt cas n°2 ici, on souhaite situer 2 Peugeot supplémentaires (même s'il y a déjà la Peugeot 504 parmi les individus actifs).

Calculs pour les individus illustratifs

Description des véhicules

Modele	CYL	PUISS	LONG	LARG	POIDS	V.MAX
Peugeot 604	2664	136	472	177	1410	180
Peugeot 304 S	1288	74	414	157	915	160

Moyenne	1631.667	84.611	433.500	166.667	1078.833	158.278
Ecart-type	363.394	19.802	21.484	5.164	133.099	11.798

Moyennes et écarts-type calculés sur l'échantillon d'apprentissage (individus actifs, n = 18).

Modele	CYL	PUISS	LONG	LARG	POIDS	V.MAX
Peugeot 604	2.8408	2.5951	1.7920	2.0010	2.4881	1.8411
Peugeot 304 S	-0.9457	-0.5359	-0.9076	-1.8719	-1.2309	0.1460

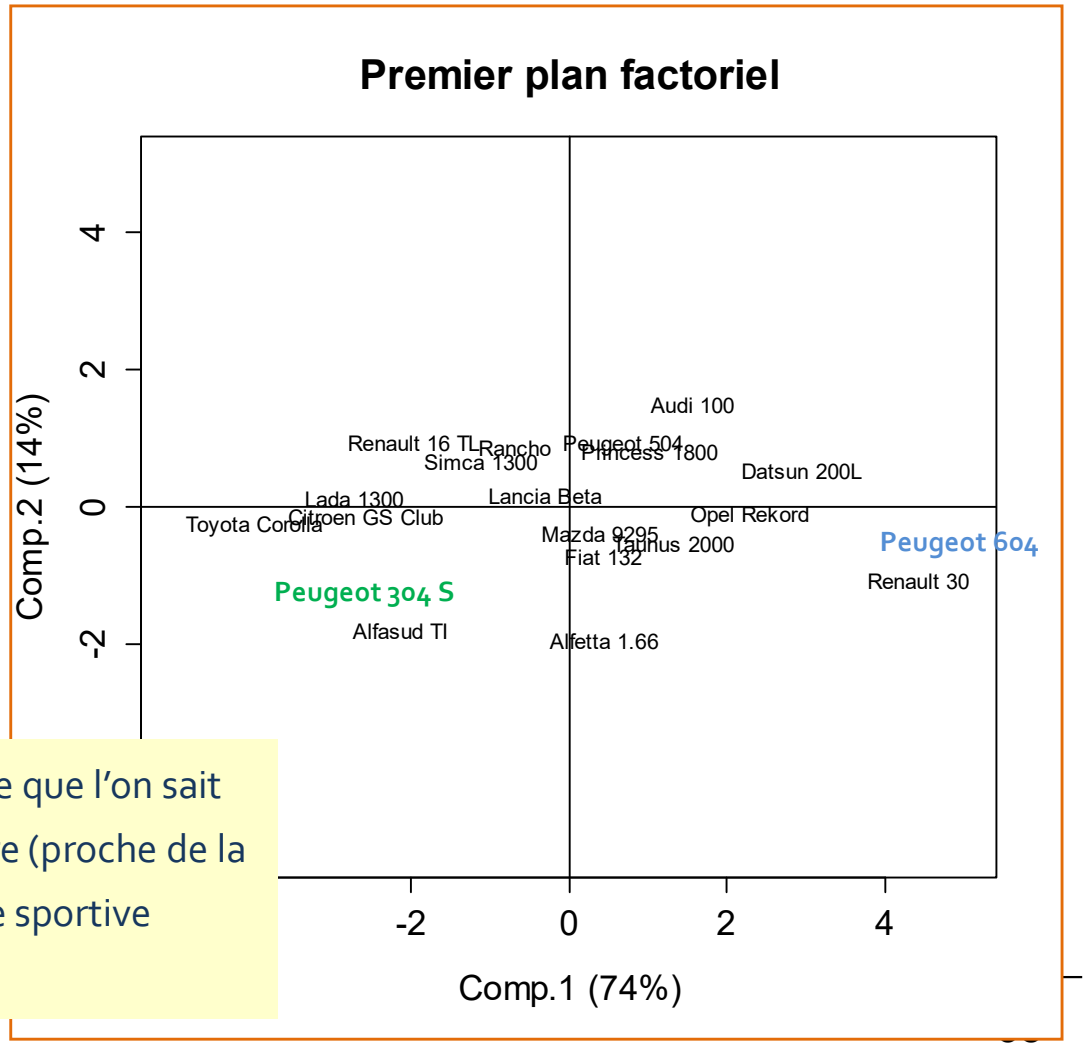
Description après centrage-réduction

Attribute	Comp.1	Comp.2
CYL	0.424936	-0.1241911
PUISS	0.4217944	-0.4157739
LONG	0.4214599	0.4118177
LARG	0.3869222	0.446087
POIDS	0.430512	0.2426758
V.MAX	0.3589443	-0.6198626

Coefficients des fonctions de projection = vecteurs propres issus de l'ACP

Modele	Comp.1	Comp.2
Peugeot 604	5.5633	-0.3386
Peugeot 304 S	-2.2122	-1.2578

Coordonnées factorielles des individus illustratifs : produit scalaire entre description (c.r.) et vecteurs propres.



Les positionnements confirment ce que l'on sait de ces véhicules : « 604 », statutaire (proche de la Renault 30); « 304 S », plutôt petite sportive (proche de l'Alfasud)