Année Universitaire 2020-2021

Analyse de données

Zouaoui Slim
zou_slim@yahoo.fr

Introduction

Définition

C'est l'ensemble des algorithmes, méthodes et technologies inspirés de plusieurs autres disciplines, pouvant servir à remplacer ou à aider l'expert humain ou le décideur dans un domaine spécifique dans le cadre de prise de décision, et ce en fouillant dans des bases de données décisionnelles des corrélations, des associations, des comportements homogènes, des formules de lien entre indicateurs, des spécification par rapport à une thématique bien déterminée, etc.

Exercice *Trace, déterminant, Valeurs et vecteurs propres*

Soit la matrice suivante :

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & -2 \\ 0 & -2 & 1 \end{pmatrix}$$

1. Calculer la trace et le déterminant de la matrice.
2. Trouver les valeurs et vecteurs propres de la matrice.
3. Déduire les deux matrices D et P telque $A = P * D * P^{-1}$

Correction exercice

Soit la matrice $A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & -2 \\ 0 & -2 & 1 \end{pmatrix}$

1- $tr(A) = \sum_{i=1}^{3} a_{ii} = 3$

$$\det(A) = \begin{vmatrix} 1 & 0 & 0 \\ 0 & 1 & -2 \\ 0 & -2 & 1 \end{vmatrix} = 1 \cdot \begin{vmatrix} 1 & -2 \\ -2 & 1 \end{vmatrix} = 1 * (1-4) = -3$$

2- Les valeurs propres de A sont déterminées à partir du polynome suivant:

$$\det(A - \lambda I) = 0 \Leftrightarrow \begin{vmatrix} 1-\lambda & 0 & 0 \\ 0 & 1-\lambda & -2 \\ 0 & -2 & 1-\lambda \end{vmatrix} = (1-\lambda) \cdot \begin{vmatrix} 1-\lambda & -2 \\ -2 & 1-\lambda \end{vmatrix}$$

$$= (1-\lambda)\left((1-\lambda)^2 - 4\right) = 0$$

$$= (1-\lambda)(1-\lambda-2)(1-\lambda+2) = 0$$

$$= (1-\lambda)(-\lambda-1)(-\lambda+3) = 0$$

$\lambda_1 = 1$; $\lambda_2 = -1$; $\lambda_3 = 3$

• U_1 un vecteur propre de A associé à la valeur propre λ_1 ssi

$AU_1 = \lambda_1 U_1$ on pose $U_1 : \begin{pmatrix} x_1 \\ y_1 \\ z_1 \end{pmatrix}$

$$\begin{cases} x_1 = x_1 \\ y_1 - 2z_1 = y_1 \\ -2y_1 + z_1 = z_1 \end{cases} \Leftrightarrow \begin{cases} x_1 = x_1 \\ z_1 = 0 \\ y_1 = 0 \end{cases} \quad U_1 : \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

$\|U_1\| = 1$

• U_2 un vecteur propre de A associé à la valeur propre λ_2 ssi

$AU_2 = \lambda_2 U_2$ on pose $U_2 : \begin{pmatrix} x_2 \\ y_2 \\ z_2 \end{pmatrix}$

$$\begin{cases} x_2 = -x_2 \\ y_2 - 2z_2 = -y_2 \\ -2y_2 + z_2 = -z_2 \end{cases} \Leftrightarrow \begin{cases} x_2 = 0 \\ y_2 = z_2 \\ y_2 = z_2 \end{cases} \quad U_2 : \begin{pmatrix} 0 \\ 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}$$

$\|U_2\| = 1$

• U_3 un vecteur propre de A associé à la valeur propre λ_3 ssi

$AU_3 = \lambda_3 U_3$ on pose $U_3 : \begin{pmatrix} x_3 \\ y_3 \\ z_3 \end{pmatrix}$

$$\begin{cases} x_3 = 3x_3 \\ y_3 - 2z_3 = 3y_3 \\ -2y_3 + z_3 = 3z_3 \end{cases} \Leftrightarrow \begin{cases} x_3 = 0 \\ y_3 = -z_3 \\ y_3 = -z_3 \end{cases} \Leftrightarrow U_3 : \begin{pmatrix} 0 \\ 1/\sqrt{2} \\ -1/\sqrt{2} \end{pmatrix}$$

$\|U_3\| = 1$

3- $A = P \cdot D \cdot P^{-1}$ avec $D : \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 3 \end{pmatrix}$ et $P : \begin{array}{c} \begin{matrix} U_1 & U_2 & U_3 \end{matrix} \\ \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1/\sqrt{2} & 1/\sqrt{2} \\ 0 & 1/\sqrt{2} & -1/\sqrt{2} \end{pmatrix} \end{array}$

Deux familles de techniques

- **Méthodes Descriptives**
  - Analyse en Composantes Principales ACP
  - Méthodes des Centres Mobiles
    K-means
  - Classification Ascendante Hiérarchique
- **Méthodes Prédictives**
  - Arbres de Décisions
  - Analyse Discriminante
  - Régression Linéaire
  - Régression Logistique
  - Réseaux de Neurones

Sommaire

- Analyse en composantes principales
- Analyse factorielle de correspondance
- Méthodes de classification
- modélisation linéaire simple et multiple

Coheris SPAD

cran.r-project.org

anaconda.com

Analyse en composantes principales

L'Analyse en Composantes Principales (**ACP**) a pour objectif de résumer un ensemble de données quantitatives. Ces données sont relatives à un grand nombre d'individus et /ou de variables illustrés dans un tableau à **n** lignes (chaque ligne représente un individu de l'échantillon étudié composé de **n** observations) et **p** colonnes. **p** étant le nombre de variables quantitatives mesurées sur les **n** individus. Elle permet notamment de :

- Décrire et représenter le réseau d'interaction entre les variables.

- Décrire et représenter les ressemblances entre les individus % à l'ensemble des variables.

Ainsi par le biais de l'**ACP**, on va synthétiser l'information fournie par les **p** variables et ce en construisant un certain nombre de variables nouvelles, les composantes principales, qui sont des combinaisons linéaires des différentes variables initiales.

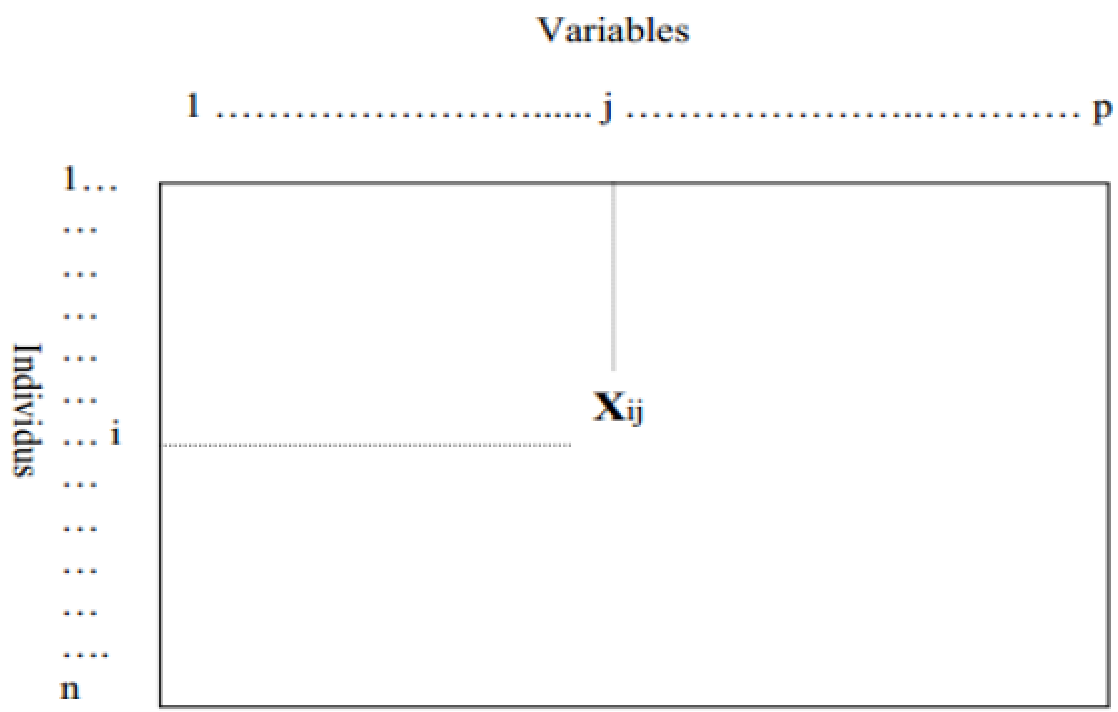
I - Présentation des données :

Soit ***X*** *un tableau à* ***n*** *lignes et* ***p*** *colonnes.*

	X_1	X_2	…..	X_{p-1}	X_p
1	x_{11}	x_{21}	…..	$x_{(p-1)1}$	x_{p1}
2	x_{12}	x_{22}	…..	$x_{(p-1)2}$	x_{p2}
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
n-1	$x_{1(n-1)}$	$x_{2(n-1)}$	…..	$x_{(p-1)(n-1)}$	$x_{p(n-1)}$
n	x_{1n}	x_{2n}	…..	$x_{(p-1)n}$	x_{pn}

On possède un tableau rectangulaire dont :

- ✓ **les colonnes sont des variables quantitatives** (mensurations, taux,…)
- ✓ **les lignes représentent des individus statistiques** (unités élémentaires telles que des êtres humains, des pays, des années…)



X : Tableau de données

Xij : Valeur de la ième observation pour la jième variable

Xi. : ième observation du tableau

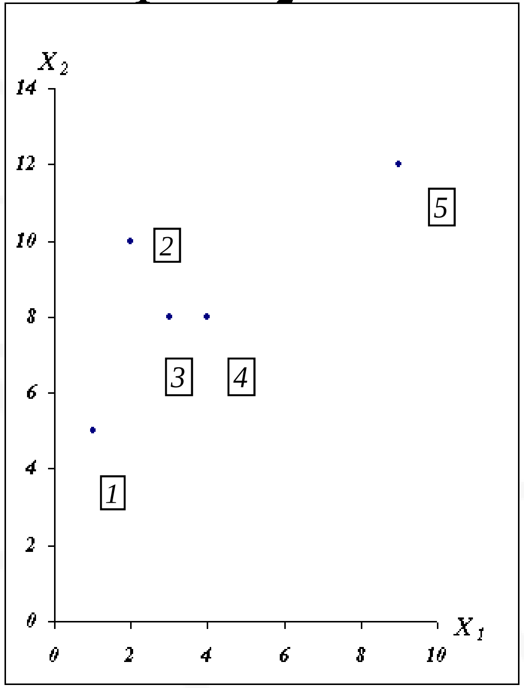
X.j : jième variable du tableau

n : effectif des individus

p : nombre de variables

Exemple : *considérons deux variables* $\boldsymbol{X_1}$ *et* $\boldsymbol{X_2}$ *mesurées sur cinq individus.*

Individu	X_1	X_2
1	1	5
2	2	10
3	3	8
4	4	8
5	9	12



$$\bar{X}_1 = 3{,}8 \quad ; \; \bar{X}_2 = 8{,}6$$

$$\sigma_{X_1} = 2{,}79 \; ; \; \sigma_{X_2} = 2{,}33$$

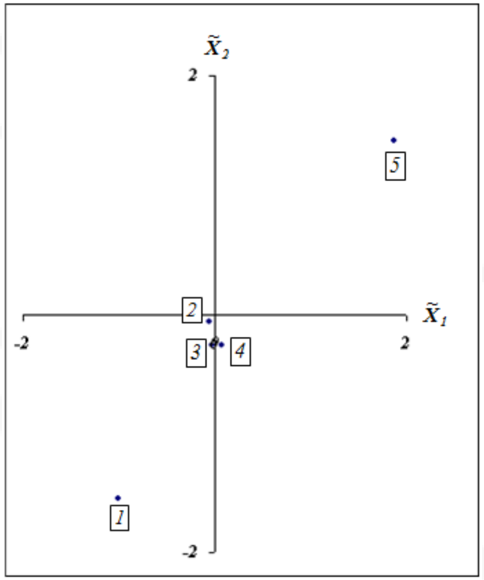
Les deux variables étudiées ne sont pas homogènes et de plus, elles n'ont pas d'importances égales. Afin de remédier à ce problème, on procède à une transformation des données et ce en déterminant les variables centrées et réduites et telles que

$$\tilde{X}_1 = \frac{X_1 - \bar{X}_1}{\sigma_{X_1}} \qquad \text{et} \qquad \tilde{X}_2 = \frac{X_2 - \bar{X}_2}{\sigma_{X_2}}$$

Ainsi, les deux variables ont la même moyenne (égale à 0) et la même variance (égale à1)

On obtient le tableau suivant :

Individu	$\tilde{X}_1$	$\tilde{X}_2$
1	-1,005	-1,543
2	-0.0646	-0,06
3	-0,0287	-0,257
4	0,072	-0,257
5	1,867	1,458



Les individus seront représentés dans un nouveau repère dont l’origine est le centre de gravité du nuage de points. On a procédé donc à une translation de l’origine du repère passant du point o au centre de gravité du nuage dont les coordonnées dans l’ancien repère étaient les moyennes respectives des variables $\mathbf{X_1}$ et $\mathbf{X_2}$.

Centrage et réduction des données

Centrer les données ne modifie pas la forme du nuage :

c'est translater !

- On peut toujours procéder par un centrage de l'entrepôt

Réduire les données :

Indispensable si les unités de mesure sont différentes !

- $x_{ij} = \frac{x_{ij} - \overline{x_k}}{s_k}$
- Permet de comparer les valeurs prises par les variables

II – La détermination des facteurs et des composantes principales :

II . 1– Covariance et coefficient de corrélation

On définit la covariance entre deux variable X et Y par :

$$\mathrm{cov}(X,Y) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})$$

Si cov (,) = 0 : les deux variables ne sont pas corrélées,
Si cov (,) > 0 : dépendance linéaire positive.
Si cov (,) < 0 : dépendance linéaire négative.

→ L'appréciation de la liaison entre deux variables et la comparaison des couples de variables n'est pas évidente avec la covariance.

On définit la matrice de variance covariance entre les variables

$$V = \begin{bmatrix} v_1^1 & \mathrm{cov}_1^2 & \dots & \mathrm{cov}_1^p \\ \mathrm{cov}_2^1 & v_2^2 & & \\ \vdots & & \ddots & \vdots \\ \mathrm{cov}_p^1 & & \dots & v_p^p \end{bmatrix}$$

On définit le coefficient de corrélation linaire entre deux variable X et Y par :

$$cor(X,Y) = \rho_{xy} = \frac{\mathrm{cov}(X,Y)}{\sqrt{\mathrm{var}(X)} * \sqrt{\mathrm{var}(Y)}} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \overline{y})^2}}$$

On a toujours (inégalité de Cauchy Schwarz) :

$$-1 \le cor(X,Y) \le 1$$

On définit la matrice de corrélation entre les variables

$$Cor = \begin{bmatrix} 1 & \rho_1^2 & \cdots & \rho_1^p \\ \rho_2^1 & 1 & & \\ \vdots & & \ddots & \vdots \\ \rho_p^1 & & \cdots & 1 \end{bmatrix}$$

Cette matrice est symétrique définit et positive

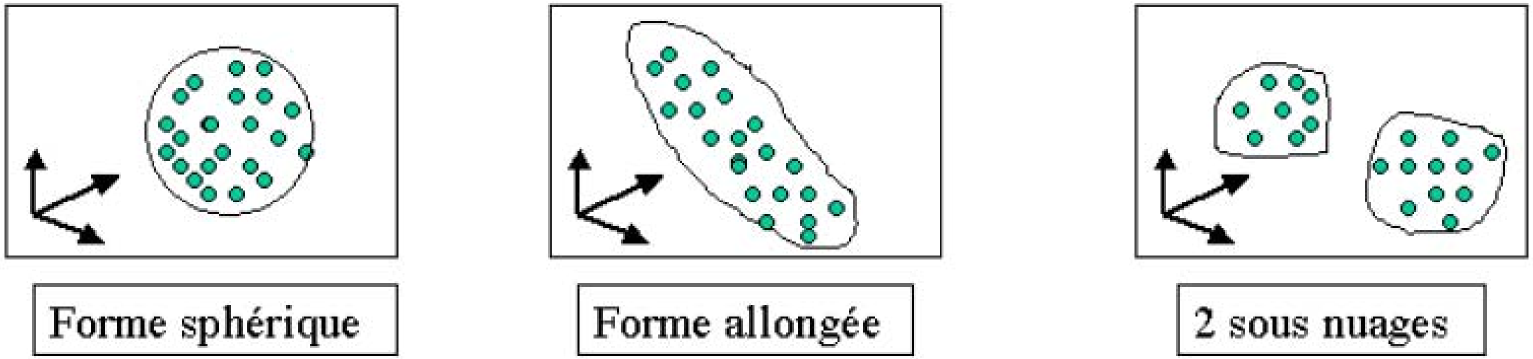
Remarque

On pose la matrice des données centrées et réduites

$$\tilde{X} = \begin{pmatrix} \tilde{x}_{11} & \tilde{x}_{21} & \tilde{x}_{31} & \cdots & \tilde{x}_{p1} \\ \tilde{x}_{12} & \tilde{x}_{22} & \tilde{x}_{32} & \cdots & \tilde{x}_{p2} \\ \tilde{x}_{13} & \tilde{x}_{23} & \tilde{x}_{33} & \cdots & \tilde{x}_{p3} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \tilde{x}_{1n} & \tilde{x}_{2n} & \tilde{x}_{3n} & \cdots & \tilde{x}_{pn} \end{pmatrix} \qquad \tilde{x}_{ij} = \frac{x_{ij} - \overline{x}_j}{\sigma_{X_j}}$$

$$Cor = \frac{1}{n-1} tr(\tilde{X})^*(\tilde{X})$$

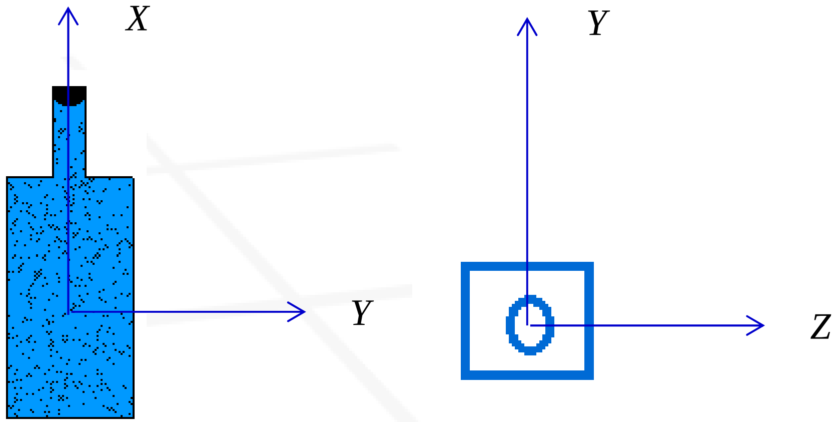
L'examen de la forme de nuage de points peut présenter un grand intérêt pour l'étude de la structure de données et pour les relations qui existent entre les différents points.



Un tel examen n'est plus possible si la dimension de l'espace considéré est élevé (>3). La solution est l'analyse factorielle qui permet une visualisation de ce nuage de points en le projetant sur un espace de dimension réduite.

II . 2– La détermination des composantes principales :

Si on vous demande de représenter une bouteille que dessinerez-vous ?



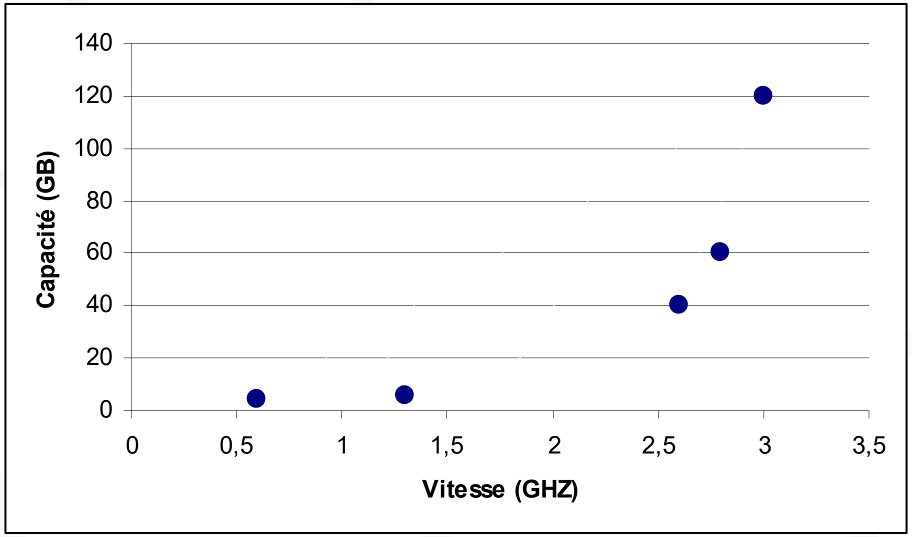
Cette partie consiste à synthétiser les données contenues dans le tableau des données transformées (centrées et réduites). Pour cela, on construit un nombre de facteurs nouvelles $\mathbf{C_1}$, $\mathbf{C_2}$, $\mathbf{C_3}$, …et $\mathbf{C_p}$ appelées composantes principales, permettant de saisir l'essentiel du tableau .

II . 2– La détermination des composantes principales :

II.2.1– Analyse de l'espace des individus

Exemple : *considérons deux variables vitesse d'un microprocesseur et capacité d'une disque dur mesurées sur cinq ordinateurs. Dans ce cas, on peut représenter les cinq ordinateurs dans le plan formé par les deux variables **VIT et CAP.***

Individu	***VIT (GHZ)***	***CAP (GB)***
1	*1.3*	*6*
2	*2.6*	*40*
3	*3*	*120*
4	*0.6*	*4*
5	*2.8*	*60*



On a : $\overline{VIT} = 2.06$ $\overline{CAP} = 46$

$\sigma_{VIT} = 1.05$ $\sigma_{CAP} = 47.06$

On constate que les deux variables n'ont pas le même ordre de grandeur. La différence d'une unité de mesure sur la vitesse n'a pas la même importance que celle d'une unité de mesure sur la capacité.

II . 2– La détermination des composantes principales :

II.2.1– Analyse de l'espace des individus

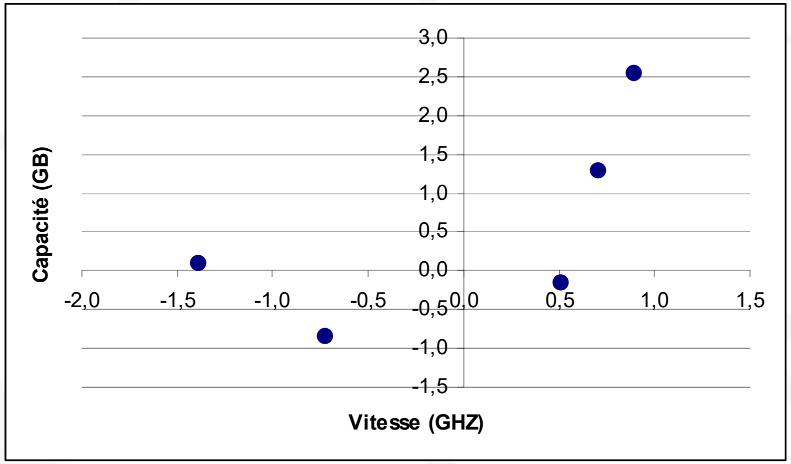
;

on procède à une transformation des données et ce en déterminant les variables centrées et réduites et on obtient le tableau $\widetilde{\mathbf{X}}$

$$\widetilde{VIT} = \frac{VIT - \overline{VIT}}{\sigma_{VIT}} \qquad \widetilde{CAP} = \frac{CAP - \overline{CAP}}{\sigma_{CAP}}$$

$\widetilde{\mathbf{X}}$

Individu	$\widetilde{VIT}$	$\widetilde{CAP}$
1	-0,724	*6*
2	0,514	*40*
3	0,895	*120*
4	-1,390	*4*
5	0,705	*60*



La seconde étape consiste à synthétiser les données contenues dans le tableau des données transformées (centrées et réduites). Pour cela, on construit un nombre de variables nouvelles C_1, C_2, C_3, *…et* C_p *appelées composantes principales, permettant de saisir l'essentiel du tableau* $\widetilde{\mathbf{X}}$

II . 2– La détermination des composantes principales :

II.2.1– Analyse de l'espace des individus

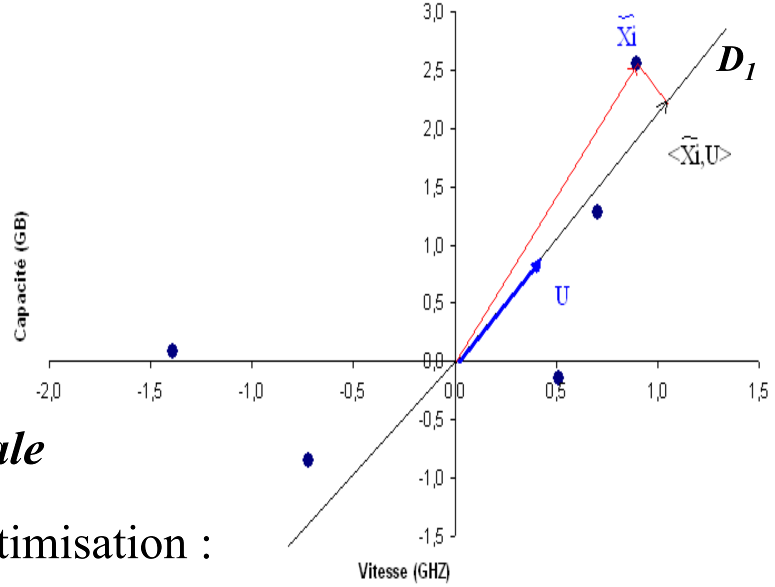
Dans cet espace, les n individus forment un nuage de points. L'objet de l'Analyse en Composantes Principales est de décrire de façon synthétique la dispersion du nuage de points.

A la première étape, l'ACP détermine l'axe D_1 passant par l'origine (le centre de gravité du nuage) selon lequel la dispersion du nuage de points est maximale. Cet axe D_1 passe au plus près du nuage de points, c'est-à-dire est tel que la moyenne des carrés des distances entre les n points et l'axe D_1 est minimale. Soit u_1 le vecteur directeur normé de D_1. u_1 est alors le vecteur propre normé associé à la valeur propre la plus élevée de la matrice de corrélation entre les variables.

Démonstration

Cet axe D_1 passe au plus près du nuage de points, c'est-à-dire est tel que l'inertie entre les n points et l'axe D_1 est minimale.

Soit U le vecteur directeur normé de D_1



➡ ***Inertie Minimale donc Inertie expliquée maximale***

La recherche d'un maximum devient un problème d'optimisation :

$$Max\ (U'\tilde{X}'\tilde{X}U) \quad sous\ contrainte\ \|U\| = 1$$

Soit le lagrangien définit par : $L = \frac{1}{n-1}(U'\tilde{X}'\tilde{X}U) - \lambda(U'U - 1)$

$$\frac{\partial L}{\partial U} = \frac{\partial}{\partial U}\left[\frac{1}{n-1}(U'\tilde{X}'\tilde{X}U) - \lambda(U'U - 1)\right] = 0$$

$$\frac{2}{n-1}\tilde{X}'\tilde{X}U - 2\lambda U = 0 \iff \frac{1}{n-1}\tilde{X}'\tilde{X}U = \lambda U$$

➡ **Donc U est un vecteur propre de la matrice $\frac{1}{n-1}\tilde{X}'\tilde{X}$ qui est égale à la matrice de corrélation entre les variables associé à la valeur propre λ**

II . 2– La détermination des composantes principales :

Ainsi à la première étape, l'ACP fournit la meilleure représentation unidimensionnelle possible du nuage de point mais elle s'avère insuffisante pour décrire complètement le nuage de n points. La dispersion du nuage dans les directions de l'espace orthogonales à D_1 n'est pas décrite par cette étape.

Durant la seconde étape, l'ACP détermine un axe D_2 de vecteur directeur normé u_2 orthogonal à u_1 passant au plus près du nuage de points. Le vecteur u_2 est le vecteur propre normé de la matrice de corrélation associé à sa deuxième valeur propre.

On continue ainsi de suite la procédure, afin de compléter la description du nuage de points donnée par les deux premières étapes.

A l'étape k, l'ACP détermine l'axe D_k passant par l'origine, de vecteur directeur normé u_k orthogonal aux différents vecteurs u_l ($l=1, 2, 3,\ldots, k-1$) selon lequel la dispersion du nuage de points est maximale. Cet axe D_k passe au plus près du nuage de points.

ACP fournit une image simplifiée de *N'*

- La plus fidèle possible
- À travers le sous-espace qui résume le mieux les données

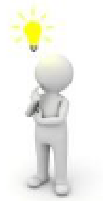
La qualité de l'image de *N'*

- Restitue fidèlement la forme générale du nuage
- Meilleur représentation de la diversité et de la variabilité

La quantification de la qualité de l'image :

Inertie totale = variance généralisée à plusieurs dimensions

Diagonalisation de données sous forme matricielle



II.2.2– Contributions des composantes principales à l'inertie totale :

En utilisant le théorème de Huygens, on peut décomposer l'inertie totale du nuage des individus

$$I_G = I_{\Delta_1^*} \otimes I_{\Delta_1^*} \otimes ....\otimes I_{\Delta_p^*} = \lambda_1 + \lambda_1 + ...+ \lambda_p$$

La contribution absolue de l'axe Δ_k à l'inertie totale du nuage des individus est égale à :

$$Contributions(\Delta_k / I_G) = \lambda_k$$ Valeur propre associé

Sa contribution relative est égale à :

$$cr(\Delta_k / I_G) = \frac{\lambda_k}{\lambda_1 + \lambda_2 + ... + \lambda_p}$$

On emploie souvent l'expression « le pourcentage d'inertie ou d'information »

On peut étendre ces définitions à tous les sous-espaces engendrés par les nouveaux axes. Ainsi, le pourcentage d'inertie expliqué par le plan engendré par les deux premiers axes Δ_1 et Δ_2 est égal à :

$$cr(\Delta_1 \otimes \Delta_2 / I_G) = \frac{\lambda_1 + \lambda_2}{\lambda_1 + \lambda_2 + ... + \lambda_p}$$

Ces pourcentages d'inertie sont des indicateurs qui rendent compte de la part de variabilité du nuage des individus expliquée par ces sous-espaces. Si les dernières valeurs propres ont des valeurs faibles, on pourra négliger la variabilité qu'expliquent les axes correspondants.

Reprenons l'exemple

Continuons l'exemple précédemment cité.

La matrice de corrélation entre les variables s'écrit :

$$\frac{1}{n}\widetilde{X}'\,\widetilde{X} = \begin{pmatrix} 1 & 0{,}84 \\ 0{,}84 & 1 \end{pmatrix}$$

Les facteurs sont les vecteurs propres normés de cette matrice de corrélation.

Le premier facteur associé à la valeur propre **1,84** est égal à $\begin{pmatrix} 0{,}707 \\ 0{,}707 \end{pmatrix}$

tandis que le second facteur associé à la deuxième valeur propre **0,16** est égal à $\begin{pmatrix} 0{,}707 \\ -0{,}707 \end{pmatrix}$

Par conséquent, les composantes principales s'écrivent :

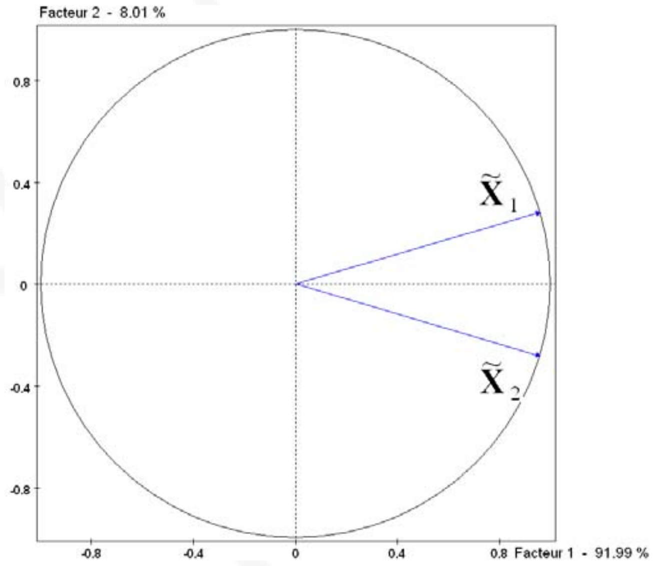
$$C_1 = 0{,}707\,\tilde{X}_1 + 0{,}707\,\tilde{X}_2 \qquad C_2 = 0{,}707\,\tilde{X}_1 - 0{,}707\,\tilde{X}_2$$

Et les pourcentages de variance expliquée sont **(1,84 /2)** soit 91,99% pour le premier axe et **(0,16 /2)** soit 8,01% pour le second axe.

Les corrélations entre les variables de départ et les composantes principales peuvent être déterminées comme suit :

$$\begin{pmatrix} \rho_{C_1;\tilde{X}_1} \\ \rho_{C_1;\tilde{X}_2} \end{pmatrix} = \sqrt{1{,}84}\begin{pmatrix} 0{,}707 \\ 0{,}707 \end{pmatrix} = \begin{pmatrix} 0{,}96 \\ 0{,}96 \end{pmatrix} = \begin{pmatrix} Cord_{C_1}(\tilde{X}_1) \\ Cord_{C_1}(\tilde{X}_2) \end{pmatrix}$$

$$\begin{pmatrix} \rho_{C_2;\tilde{X}_1} \\ \rho_{C_2;\tilde{X}_2} \end{pmatrix} = \sqrt{0{,}16}\begin{pmatrix} 0{,}707 \\ -0{,}707 \end{pmatrix} = \begin{pmatrix} 0{,}28 \\ -0{,}28 \end{pmatrix} = \begin{pmatrix} Cord_{C_2}(\tilde{X}_1) \\ Cord_{C_2}(\tilde{X}_2) \end{pmatrix}$$



Cercle de corrélation

III– La représentation des individus :

L’information pertinente est donc celle donnée par les premières étapes. Il s’agit maintenant d’analyser simultanément les résultats de ces premières étapes. Cette analyse se fait en établissant des cartes de proximités entre les individus et des cartes de corrélation entre les variables.

Pour faire la représentation des individus dans les plans définis par les nouveaux axes, il suffit de calculer les coordonnées des individus dans les nouveaux axes.

La lecture des graphiques est facilitée par le calcul de deux aides à l’interprétation : la qualité de représentation d’un individu et les contributions des individus à la variance.

III–1 Coordonnées des individus :

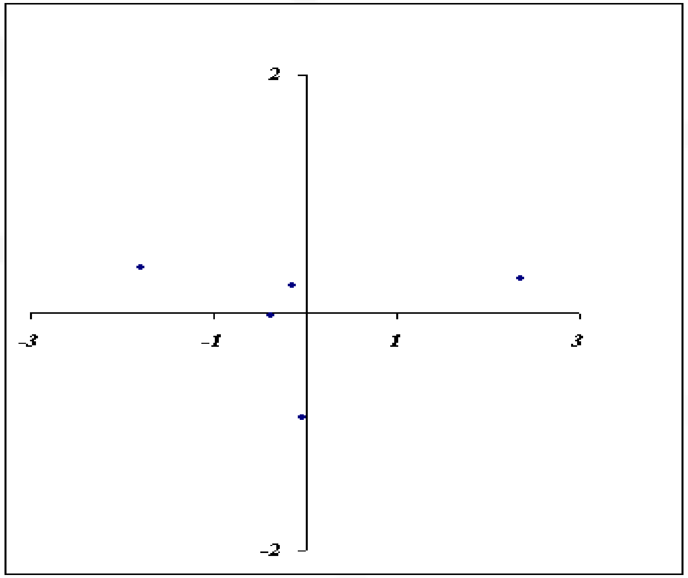
Les coordonnées de $\tilde{X}_1$ sont donc 0,96 et 0,28 tandis que celles de $\tilde{X}_2$ sont 0,96 et -0,28.

$\tilde{X}_1$ et $\tilde{X}_2$ sont positivement et fortement corrélées avec la première composante principale. Au contraire, ces deux variables sont assez faiblement corrélées avec la seconde composante principale.

Comme $C_1 = 0{,}707\,\tilde{X}_1 + 0{,}707\,\tilde{X}_2$ $\quad C_2 = 0{,}707\,\tilde{X}_1 - 0{,}707\,\tilde{X}_2$

on obtient les coordonnées des individus.

Individu	C1	C2
1	-1,802	0,381
2	-0 ,032	-0,881
3	-0,385	-0,021
4	0,131	-0,233
5	2,351	0,289



III–2 La qualité de représentation des individus :

Dans l'espace des individus, on dispose de deux bases : la base d'origine : dans cette base les coordonnées de l'individu i sont $\tilde{\mathbf{x}}_{i1}$ $\tilde{\mathbf{x}}_{i2}$,… $\tilde{\mathbf{x}}_{ip}$ .Et la base orthonormées constituée par les p facteurs : dans cette base les coordonnées de l'individu i sont $\mathbf{y}_{i1}$, $\mathbf{y}_{i2}$, …$\mathbf{y}_{ip}$

Le carré de la distance de l'individu i au centre du nuage est égal à $\sum_{j=1}^{p} \tilde{\mathbf{x}}_{ij}^2$ en effectuant les calculs avec la première base et $\sum_{j=1}^{p} \mathbf{y}_{ij}^2$ en effectuant les calculs avec la seconde base et donc $\sum_{j=1}^{p} \tilde{\mathbf{x}}_{ij}^2 = \sum_{j=1}^{p} \mathbf{y}_{ij}^2$

La qualité de représentation de l'individu i sur l'axe Δ_k est donc mesurée par $\dfrac{y_{ik}^2}{\sum_{j=1}^{p} \tilde{x}_{ij}^2} = \dfrac{y_{ik}^2}{\sum_{j=1}^{p} y_{ij}^2}$

Cette qualité de représentation est égale au carré du cosinus de l'angle entre le vecteur représentatif du point i et le vecteur directeur de l'axe Δ_k.

III–3 Les contributions des individus à la variance :

La variance expliquée à l'étape r est égale à la valeur propre λ_r

qui est égale à $\frac{1}{n}\sum_{i=1}^{n} c_{ri}^2$

La part de cette variance due à l'individu i est : $\frac{1}{n} c_{ri}^2$

La contribution de l'individu i à la variance de l'axe r est donc mesurée par :

$$\frac{\frac{1}{n} c_{ri}^2}{\frac{1}{n}\sum_{i=1}^{n} c_{ri}^2} = \frac{\frac{1}{n} c_{ri}^2}{\lambda_r} = \frac{c_{ri}^2}{n\lambda_r}$$

Pour un axe donné la somme des contributions de tous les individus est égale à 100%. Si la contribution d'un individu à un axe donné est importante, ceci signifie que cet individu joue un rôle important dans la construction de cet axe.

Reprenons l'exemple

Pour faciliter l'interprétation des résultats calculons maintenant, la qualité de représentation de chaque individu pour chacun des deux axes, à partir des coordonnées de chaque individu sur ces deux axes.

Pour l'individu 4, par exemple, la qualité de représentation est :

Axe1 : $$\frac{(-0{,}131)^2}{(0{,}072)^2+(-0{,}257)^2}=0{,}24$$

Axe 2 : $$\frac{(0{,}233)^2}{(0{,}072)^2+(-0{,}257)^2}=0{,}76$$

Pour ce même individu 4, la contribution à la variance est :

Axe1 : $$\frac{(-0{,}131)^2}{5(1{,}788)}=0{,}002$$

Axe 2 : $$\frac{(0{,}233)^2}{5(0{,}212)}=0{,}051$$

Et on obtient pour l'ensemble des individus les tableaux suivants :

Qualité de représentation			
Individu	***Axe1***	***Axe 2***	***Total***
1	*0,96*	*0,04*	***1***
2	*0,01*	*0,99*	***1***
3	*0,99*	*0,01*	***1***
4	*0,24*	*0,76*	***1***
5	*0,98*	*0,02*	***1***

Contribution à la variance		
Individu	***Axe1***	***Axe 2***
1	*0,36*	*0,14*
2	*0,00*	*0,73*
3	*0,02*	*0,00*
4	*0,00*	*0,05*
5	*0,62*	*0,08*
Total	*1*	*1*

Les individus 1 et 5 sont ceux qui contribuent le plus fortement à la variance sur le premier axe. Sur cet axe, ils s'opposent, puisque l'individu 1 est à gauche de l'axe et l'individu 5 à la droite de l'axe. L'interprétation est simple : le premier axe est lié fortement aux variables $\tilde{X}_1$ *et* $\tilde{X}_2$ *et représente « la taille » des individus : les individus 1 et 5 sont ceux qui connaissent des valeurs extrêmes à la fois pour* $\tilde{X}_1$ *et* $\tilde{X}_2$ *, petites pour l'individu 1 et grandes pour l'individu 5.*

Notons que l'individu 3 est presque parfaitement représenté sur l'axe 1 : sa position correspond aux valeurs qu'il prend pour variables $\tilde{X}_1$ et $\tilde{X}_2$ c'est-à-dire légèrement en dessous de la moyenne pour chacune des deux variables.

C'est l'individu 2 qui contribue à la variance de l'axe 2 ; en fait, cet axe est lié positivement à $\tilde{X}_1$, et négativement à $\tilde{X}_2$ et la position de l'individu 2 est due à la faible valeur qu'il prend pour $\tilde{X}_1$ par rapport à la forte valeur prise par $\tilde{X}_2$. A l'opposé, l'individu 4, bien représenté sur le second axe, doit sa position à une valeur de $\tilde{X}_1$ relativement forte par rapport à la valeur prise par $\tilde{X}_2$.

INTÉRÊTS DE L'ACP

Visualisation

- *Représentation assez fidèle des individus d'une population en 2 dimensions via de nouvelles variables*

Pertinence de Variables

- *Détection des facteurs les plus pertinents dans une dynamique observée*

Relation

Variable / Variable

- *Mesure du taux de dépendance entre les variables*

OBJECTIFS – ÉTUDES DES variables

Ressemblance

- *Réseau d'interaction entre les variables*

Informations

- *Détection des indicateurs synthétiques*

Corrélations

- *Regroupement des variables*

OBJECTIFS – ÉTUDES DES INDIVIDUS

Typologie

- *Groupes d'individus homogènes*

Ressemblance

- *Du point de vue des variables*

Différence

- *Du point de vue des variables*

Application- Analyse en Composantes Principales (ACP)

On dispose de 6 variables continues observées sur 20 voitures. On va présenter dans la suite la liste des variables :nom du véhicule, longueur (m), largeur (m), hauteur (m), puissance (chevaux), vitesse maxi (Km/h), 0 à 100 km (secondes)

A partir de ces données, il vous est demandé de faire une analyse en composantes principales :

1- A partir de la matrice des corrélations, peut-on induire des liaisons entre les variables continues.
2- En prenant en compte les corrélations variables-facteurs et le graphique du cercle des corrélations, donnez une interprétation du premier et deuxième axe factoriel.
3- Vous appuyant sur les contributions et $\cos^2$ des observations (véhicules) ainsi que la carte des individus sur le premier plan factoriel interpréter la distribution des véhicules dans ce plan factoriel.

```
> voit = read.table("voitures.txt", header=T,sep="\t")
> Fix(voit)
```

Data Editor

	Voiture	longueur	largeur	hauteur	puissance	vitesse.maxi	X0_.100km
1	Skoda Octavia	4,51	1,73	1,45	90	181	12,2
2	Renault Clio	3,77	1,64	1,42	65	161	14,8
3	Peugeot 206	3,83	1,65	1,43	90	180	11
4	Toyota Yaris	3,61	1,66	1,5	75	168	11,9
5	Citroen Picasso	4,28	1,75	1,64	90	169	13
6	Peugeot 307	4,2	1,75	1,51	107	185	11,2
7	Opel Astra	4,11	1,71	1,42	125	202	9,6
8	Seat Leon	4,18	1,74	1,46	110	190	10,6
9	Toyota Avensis	4,65	1,76	1,66	116	180	12,5
10	Renault Laguna	4,58	1,75	1,43	120	195	10,6
11	BMW 320	4,26	1,75	1,41	150	217	8,7
12	Citroen C5	4,62	1,77	1,48	110	186	12
13	Renault Espace	4,86	1,86	1,75	117	171	13,4
14	Volvo S80	4,82	1,83	1,45	130	196	10,7
15	Audi A4	4,55	1,77	1,43	163	220	9,2
16	Skoda Superb	4,8	1,77	1,47	130	206	10
17	Mercedes C30	4,53	1,73	1,41	231	250	6,8
18	Alfa Romeo 147	4,21	1,76	1,41	250	246	6,3
19	BMW Z4	4,09	1,78	1,29	231	250	5,9

> voiture=voit[,-1]
> voiture_cr=scale(voiture)
> fix(voiture_cr)

Data Editor

	longueur	largeur	hauteur	puissance	vitesse.maxi	X0.100km
1	0.383918	-0.3427665	-0.2131965	-0.8393795	-0.6342908	0.7069872
2	-1.602791	-1.966397	-0.503919	-1.303638	-1.371838	1.821028
3	-1.441706	-1.785994	-0.4070115	-0.8393795	-0.6711681	0.1928147
4	-2.032349	-1.60559	0.271341	-1.117935	-1.113697	0.5784441
5	-0.2335725	0.01804034	1.628046	-0.8393795	-1.076819	1.049769
6	-0.4483518	0.01804034	0.3682485	-0.5236837	-0.4867813	0.2785101
7	-0.6899785	-0.7035733	-0.503919	-0.1894175	0.140134	-0.4070532
8	-0.5020466	-0.1623631	-0.116289	-0.4679726	-0.3023944	0.02142385
9	0.7597818	0.1984438	1.821861	-0.3565506	-0.6711681	0.8355303
10	0.5718499	0.01804034	-0.4070115	-0.2822692	-0.1180076	0.02142385
11	-0.2872673	0.01804034	-0.6008265	0.2748411	0.6932946	-0.7926826
12	0.6792396	0.3788472	0.077526	-0.4679726	-0.4499039	0.6212918
13	1.323577	2.002478	2.694029	-0.3379802	-1.003064	1.22116
14	1.216188	1.461268	-0.2131965	-0.09656578	-0.08113021	0.06427156
15	0.4913077	0.3788472	-0.4070115	0.5162555	0.8039267	-0.5784441
16	1.162493	0.3788472	-0.0193815	-0.09656578	0.2876435	-0.2356624
17	0.4376128	-0.3427665	-0.6008265	1.779039	1.910248	-1.606789
18	-0.4215044	0.1984438	-0.6008265	2.131875	1.762738	-1.821028
19	-0.7436734	0.5592506	-1.763717	1.779039	1.910248	-1.992418

```
> install.packages(c("FactoMineR", "factoextra"))
> library("FactoMineR")
> library("factoextra")
```

Plusieurs fonctions, de différents packages, sont disponibles dans le logiciel R pour le calcul de l'ACP:

- prcomp() et princomp() [fonction de base, package stats],
- PCA() [package FactoMineR],
- dudi.pca() [package ade4],
- epPCA() [package ExPosition]

Peu importe la fonction que vous décidez d'utiliser, vous pouvez facilement extraire et visualiser les résultats de l'ACP en utilisant les fonctions R fournies dans le package factoextra.

Ici, nous utiliserons les deux packages FactoMineR (pour l'analyse) et factoextra (pour la visualisation, des données, basée sur ggplot2).

> res=PCA(voiture_cr, scale.unit = TRUE, ncp = 6, graph = TRUE)
> summary(res)
> Print(res)

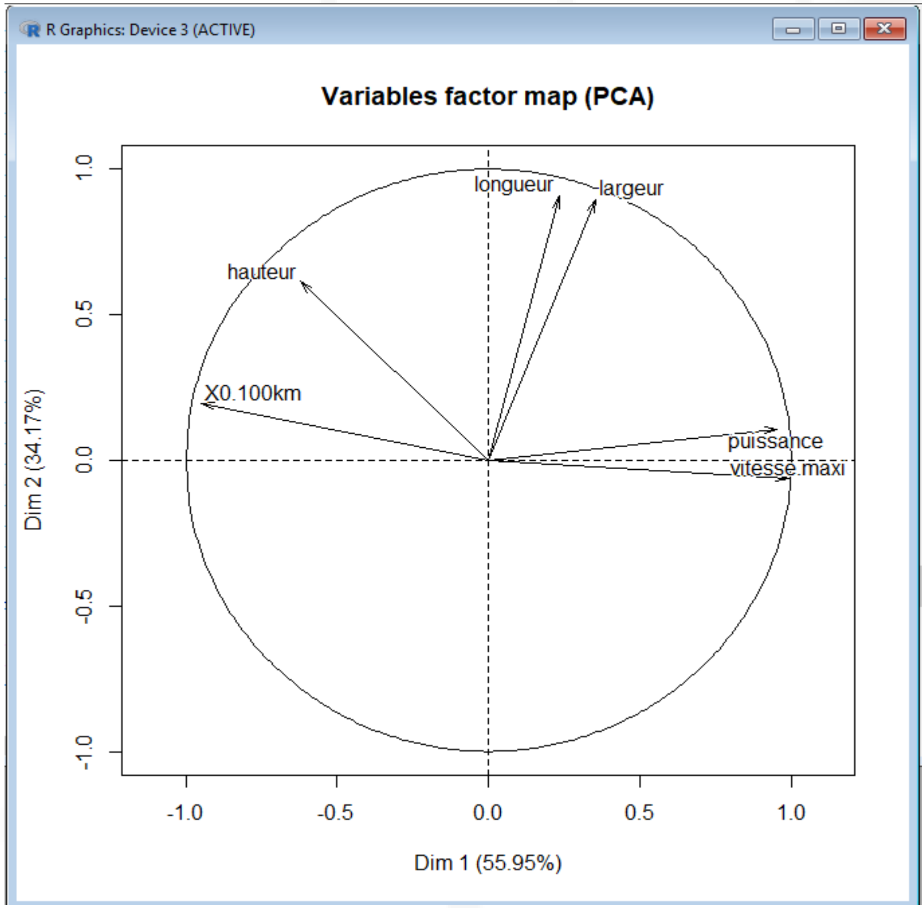
```
> res=PCA(voiture_cr, scale.unit = TRUE, ncp = 6, graph = TRUE)
> summary(res)

Call:
PCA(X = voiture_cr, scale.unit = TRUE, ncp = 6, graph = TRUE)


Eigenvalues
                       Dim.1   Dim.2   Dim.3   Dim.4   Dim.5   Dim.6
Variance               3.357   2.050   0.361   0.137   0.087   0.008
% of var.             55.948  34.166   6.020   2.275   1.453   0.138
Cumulative % of var.  55.948  90.113  96.134  98.409  99.862 100.000

Individuals (the 10 first)
                  Dist    Dim.1     ctr    cos2    Dim.2     ctr    cos2    Dim.3     ctr    cos2
1             |  1.421 | -1.119   1.866   0.621 | -0.001   0.000   0.000 | -0.837   9.693   0.347
2             |  3.782 | -2.846  12.069   0.567 | -2.307  12.979   0.372 | -0.534   3.954   0.020
3             |  2.641 | -1.322   2.602   0.250 | -2.267  12.532   0.737 |  0.067   0.061   0.001
4             |  3.180 | -2.198   7.197   0.478 | -2.185  11.641   0.472 |  0.658   5.998   0.043
5             |  2.443 | -2.197   7.193   0.809 |  0.706   1.215   0.083 |  0.799   8.832   0.107
6             |  0.987 | -0.881   1.155   0.796 | -0.098   0.023   0.010 |  0.225   0.703   0.052
7             |  1.234 |  0.140   0.029   0.013 | -1.198   3.499   0.942 |  0.010   0.001   0.000
8             |  0.797 | -0.486   0.351   0.372 | -0.501   0.612   0.395 | -0.033   0.016   0.002
9             |  2.342 | -1.504   3.368   0.412 |  1.541   5.794   0.433 |  0.724   7.249   0.095
10            |  0.786 | -0.008   0.000   0.000 |  0.191   0.089   0.059 | -0.709   6.958   0.813
```

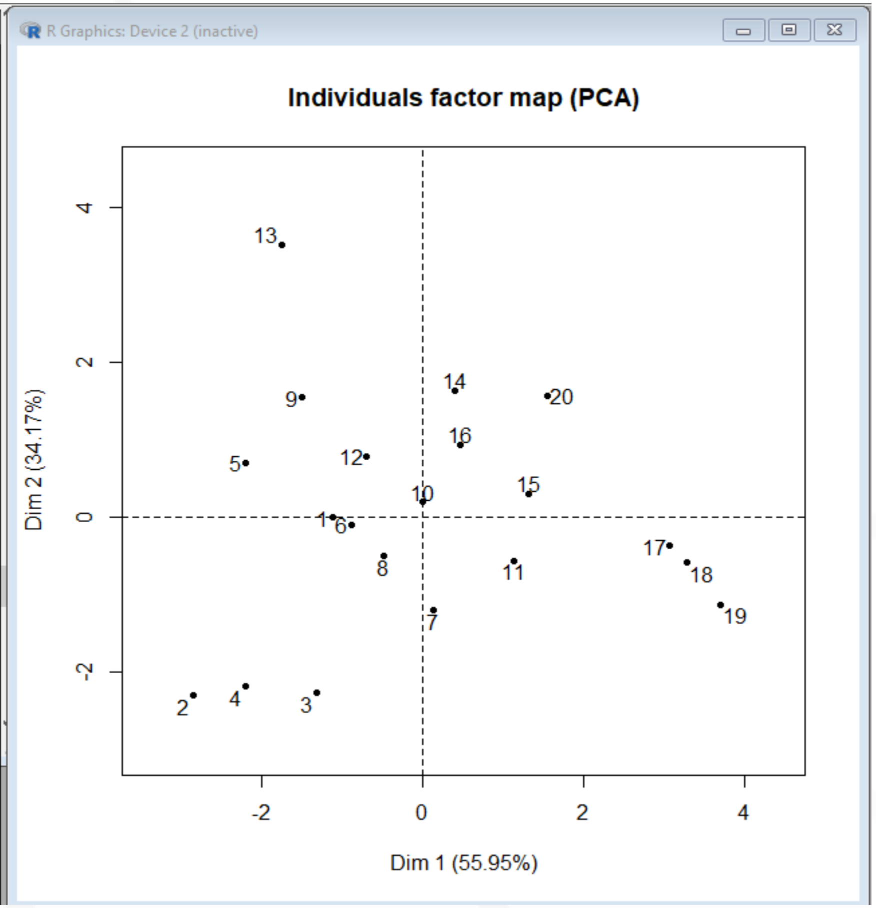
```
Variables
                 Dim.1    ctr   cos2    Dim.2    ctr   cos2    Dim.3    ctr   cos2  
longueur     |  0.232  1.603  0.054 |  0.907 40.106  0.822 | -0.277 21.315  0.077 |
largeur      |  0.353  3.721  0.125 |  0.893 38.902  0.797 | -0.033  0.294  0.001 |
hauteur      | -0.622 11.543  0.387 |  0.614 18.401  0.377 |  0.476 62.771  0.227 |
puissance    |  0.954 27.097  0.910 |  0.107  0.555  0.011 |  0.181  9.042  0.033 |
vitesse.maxi |  0.990 29.180  0.980 | -0.061  0.180  0.004 |  0.070  1.376  0.005 |
X0.100km     | -0.949 26.855  0.901 |  0.195  1.857  0.038 | -0.137  5.203  0.019 |
> res
**Results for the Principal Component Analysis (PCA)**
The analysis was performed on 20 individuals, described by 6 variables
*The results are available in the following objects:

   name                 description
1  "$eig"               "eigenvalues"
2  "$var"               "results for the variables"
3  "$var$coord"         "coord. for the variables"
4  "$var$cor"           "correlations variables - dimensions"
5  "$var$cos2"          "cos2 for the variables"
6  "$var$contrib"       "contributions of the variables"
7  "$ind"               "results for the individuals"
8  "$ind$coord"         "coord. for the individuals"
9  "$ind$cos2"          "cos2 for the individuals"
10 "$ind$contrib"       "contributions of the individuals"
11 "$call"              "summary statistics"
12 "$call$centre"       "mean of the variables"
13 "$call$ecart.type"   "standard error of the variables"
14 "$call$row.w"        "weights for the individuals"
15 "$call$col.w"        "weights for the variables"
```

Cercle de corrélation : Variables factor map (PCA)

Projection des individus dans le premier plan factoriel :

Individuals factor map (PCA)





Analyse en Composantes Principales avec Python

In [1]:
```
import os
os.chdir("C:/voitures")
# importation des donnees
import pandas
voiture = pandas.read_table("voitures.txt",sep="\t",header=0)
print(voiture.head(n=5))
```

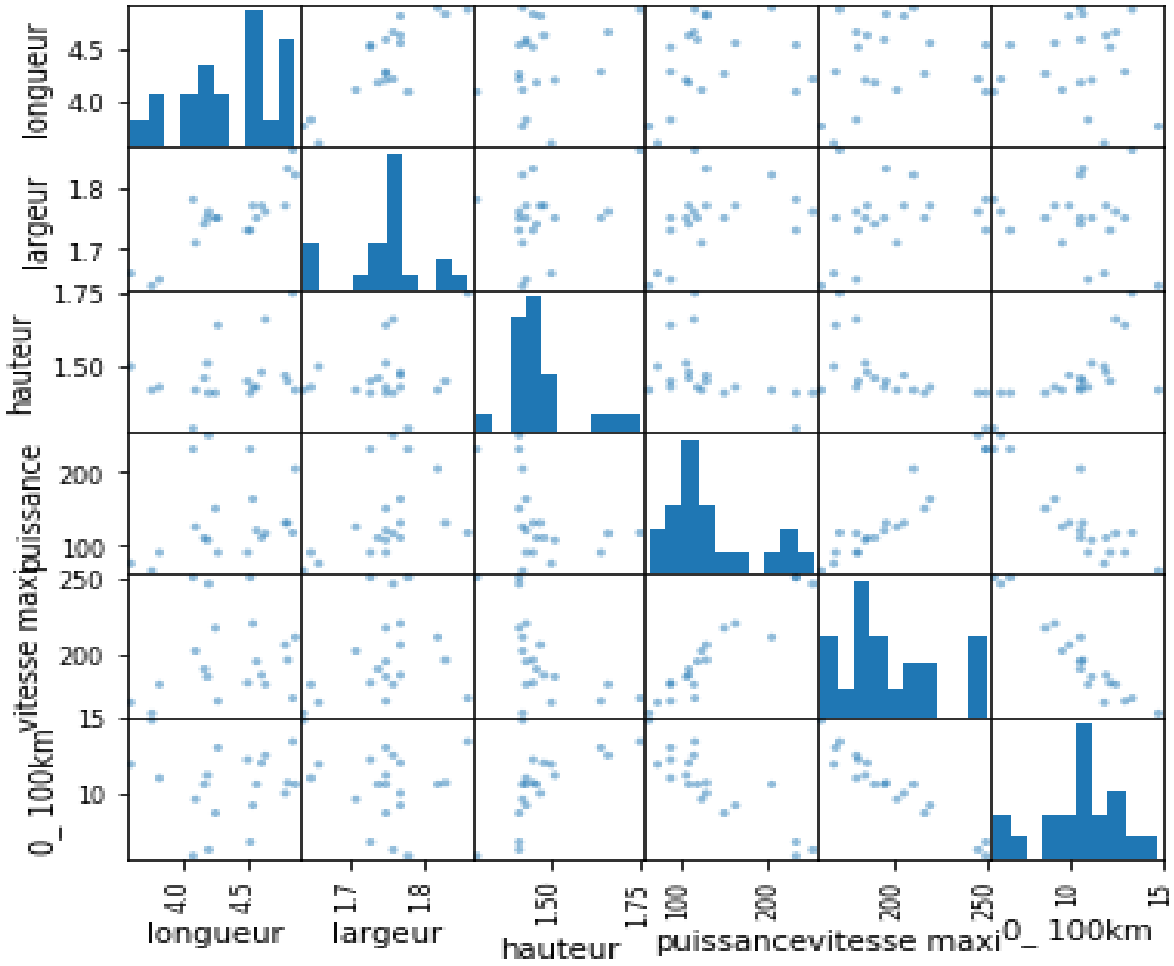
```
           Voiture  longueur  largeur  hauteur  puissance  vitesse maxi  \
0    Skoda Octavia      4.51     1.73     1.45         90           181
1    Renault Clio       3.77     1.64     1.42         65           161
2      Peugeot 206      3.83     1.65     1.43         90           180
3     Toyota Yaris      3.61     1.66     1.50         75           168
4  Citroen Picasso      4.28     1.75     1.64         90           169

   0_ 100km
0      12.2
1      14.8
2      11.0
3      11.9
4      13.0
```

In [2]:
```
# analyse descriptive des donnees
print(voiture.describe())
```

```
        longueur    largeur    hauteur   puissance  vitesse maxi   0_ 100km
count  20.000000  20.000000  20.000000   20.000000     20.000000  20.000000
mean    4.367000   1.749000   1.472000  135.200000    198.200000  10.550000
std     0.372475   0.055431   0.103191   53.849302     27.116901   2.333847
min     3.610000   1.640000   1.290000   65.000000    161.000000   5.900000
25%     4.162500   1.730000   1.420000  102.750000    180.000000   9.500000
50%     4.395000   1.750000   1.440000  118.500000    192.500000  10.650000
75%     4.627500   1.770000   1.485000  153.250000    212.500000  12.050000
max     4.880000   1.860000   1.750000  250.000000    250.000000  14.800000
```

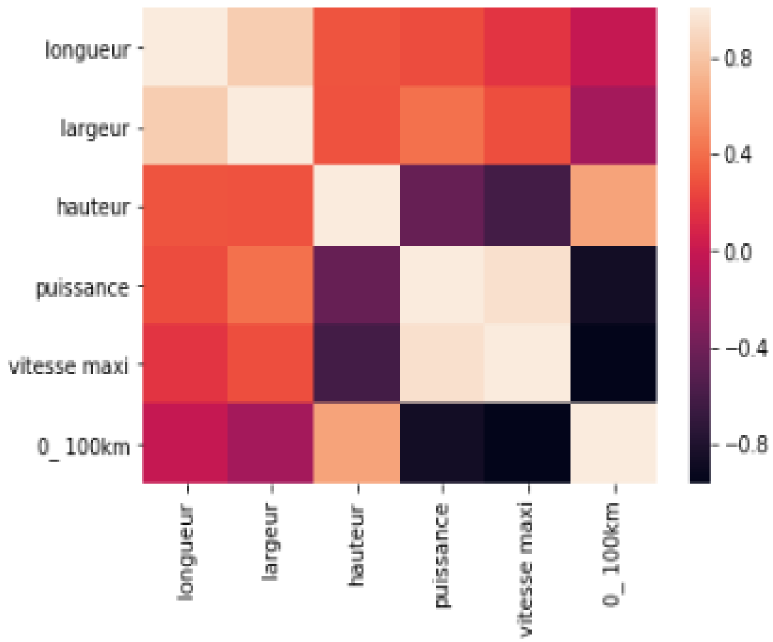
Exemple ACP sous Python

In [3]:
```
# analyse graphique bidimentionnelle
from pandas.tools.plotting import scatter_matrix
scatter_matrix(voiture,figsize=(6,6))
```



```
In [6]: # selection des variables quantitatives
        x = voiture.ix[:,1:].values
        # selection des variables qualitatives
        y = voiture.ix[:,0]
        # matrice de correlation des variables quantitatives
        corr=voiture.corr()
        print(corr)
        import seaborn as sns
        sns.heatmap(corr,
                    xticklabels=corr.columns.values,
                    yticklabels=corr.columns.values)
```

	longueur	largeur	hauteur	puissance	vitesse maxi	0_ 100km
longueur	1.000000	0.841828	0.299499	0.268942	0.171291	-0.010111
largeur	0.841828	1.000000	0.290209	0.414079	0.272905	-0.161514
hauteur	0.299499	0.290209	1.000000	-0.447420	-0.612004	0.639012
puissance	0.268942	0.414079	-0.447420	1.000000	0.945715	-0.868482
vitesse maxi	0.171291	0.272905	-0.612004	0.945715	1.000000	-0.962121
0_ 100km	-0.010111	-0.161514	0.639012	-0.868482	-0.962121	1.000000



In [7]:

```python
# variables quantitatives centrées et réduites
from sklearn import preprocessing
voiturecr = preprocessing.scale(x)
print(voiturecr)
```

```
[[ 0.39389157 -0.351671   -0.21873499 -0.86118519 -0.6507686   0.72535356]
 [-1.64442844 -2.01748099 -0.51700999 -1.33750443 -1.40747627  1.86833493]
 [-1.47915925 -1.83239099 -0.41758499 -0.86118519 -0.68860398  0.1978237 ]
 [-2.08514628 -1.64730099  0.27838999 -1.14697673 -1.14262859  0.5934711 ]
 [-0.23964033  0.018509    1.67033996 -0.86118519 -1.1047932   1.07704014]
 [-0.45999924  0.018509    0.37781499 -0.5372881  -0.49942707  0.28574534]
 [-0.70790303 -0.721851   -0.51700999 -0.19433825  0.14377446 -0.41762781]
 [-0.51508897 -0.166581   -0.11931    -0.48012979 -0.31025015  0.02198041]
 [ 0.77951968  0.203599    1.86918996 -0.36581318 -0.68860398  0.85723603]
 [ 0.58670562  0.018509   -0.41758499 -0.2896021  -0.12107323  0.02198041]
 [-0.29473005  0.018509   -0.61643499  0.28198099  0.71130521 -0.81327521]
 [ 0.69688508  0.388689    0.07954    -0.48012979 -0.46159168  0.63743192]
 [ 1.35796184  2.05449899  2.76401494 -0.34676041 -1.02912244  1.25288343]
 [ 1.24778238  1.49922899 -0.21873499 -0.0990744  -0.08323784  0.06594123]
 [ 0.50407103  0.388689   -0.41758499  0.529667    0.82481137 -0.5934711 ]
 [ 1.19269265  0.388689   -0.019885   -0.0990744   0.29511599 -0.24178452]
 [ 0.4489813  -0.351671   -0.61643499  1.82525533  1.95987288 -1.64853082]
 [-0.43245438  0.203599   -0.61643499  2.18725795  1.80853134 -1.86833493]
 [-0.76299276  0.573779   -1.80953496  1.82525533  1.95987288 -2.04417822]
 [ 1.41305157  1.31413899 -0.51700999  1.31083055  0.48429291  0.02198041]]
```

In [9]:

```python
# Déterminer les valeurs et vecteurs propres de la matrice de corrélation
import numpy as np
eig_vals, eig_vecs = np.linalg.eig(corr)
print('Eigenvectors \n%s' %eig_vecs)
print('\nEigenvalues \n%s' %eig_vals)
```

```
Eigenvectors
[[-0.12661268 -0.63328929  0.4616774   0.14650792  0.57836858 -0.11744401]
 [-0.19290792 -0.62371208  0.05422465 -0.1023808  -0.7480652  -0.02731461]
 [ 0.3397514  -0.42896434 -0.79227879 -0.06335589  0.22614171 -0.13303704]
 [-0.52055248 -0.07447686 -0.30070463  0.34436632  0.10148048  0.7100483 ]
 [-0.54018742  0.04240428 -0.11728417 -0.80159474  0.20815386 -0.08222958]
 [ 0.51821561 -0.13628126  0.22809894 -0.45044128  0.03349405  0.67589324]]

Eigenvalues
[3.35685435 2.04994817 0.36122776 0.00825088 0.13652922 0.08718961]
```

```python
In [9]: # Déterminer les valeurs et vecteurs propres de la matrice de corrélation
import numpy as np
eig_vals, eig_vecs = np.linalg.eig(corr)
print('Eigenvectors \n%s' %eig_vecs)
print('\nEigenvalues \n%s' %eig_vals)
```

```
Eigenvectors
[[-0.12661268 -0.63328929  0.4616774   0.14650792  0.57836858 -0.11744401]
 [-0.19290792 -0.62371208  0.05422465 -0.1023808  -0.7480652  -0.02731461]
 [ 0.3397514  -0.42896434 -0.79227879 -0.06335589  0.22614171 -0.13303704]
 [-0.52055248 -0.07447686 -0.30070463  0.34436632  0.10148048  0.7100483 ]
 [-0.54018742  0.04240428 -0.11728417 -0.80159474  0.20815386 -0.08222958]
 [ 0.51821561 -0.13628126  0.22809894 -0.45044128  0.03349405  0.67589324]]

Eigenvalues
[3.35685435 2.04994817 0.36122776 0.00825088 0.13652922 0.08718961]
```

TP voiture (Sortie Logiciel SPAD)

NUM . IDEN - LIBELLE	EFFECTIF	POIDS	MOYENNE	ECART-TYPE	MINIMUM	MAXIMUM
2 . C2 - longueur	20	20.00	4.37	0.36	3.61	4.88
3 . C3 - largeur	20	20.00	1.75	0.05	1.64	1.86
4 . C4 - hauteur	20	20.00	1.47	0.10	1.29	1.75
5 . C5 - puissance	20	20.00	135.20	52.49	65.00	250.00
6 . C6 - vitesse maxi	20	20.00	198.20	26.43	161.00	250.00
7 . C7 - 0 100km	20	20.00	10.55	2.27	5.90	14.80

MATRICE DES CORRELATIONS

	C2	C3	C4	C5	C6	C7
C2	1.00					
C3	0.84	1.00				
C4	0.30	0.29	1.00			
C5	0.27	0.41	-0.45	1.00		
C6	0.17	0.27	-0.61	0.95	1.00	
C7	-0.01	-0.16	0.64	-0.87	-0.96	1.00

Analyse en Composantes Principales avec SPAD

VALEURS PROPRES

APERCU DE LA PRECISION DES CALCULS : TRACE AVANT DIAGONALISATION .. 6.0000

SOMME DES VALEURS PROPRES 6.0000

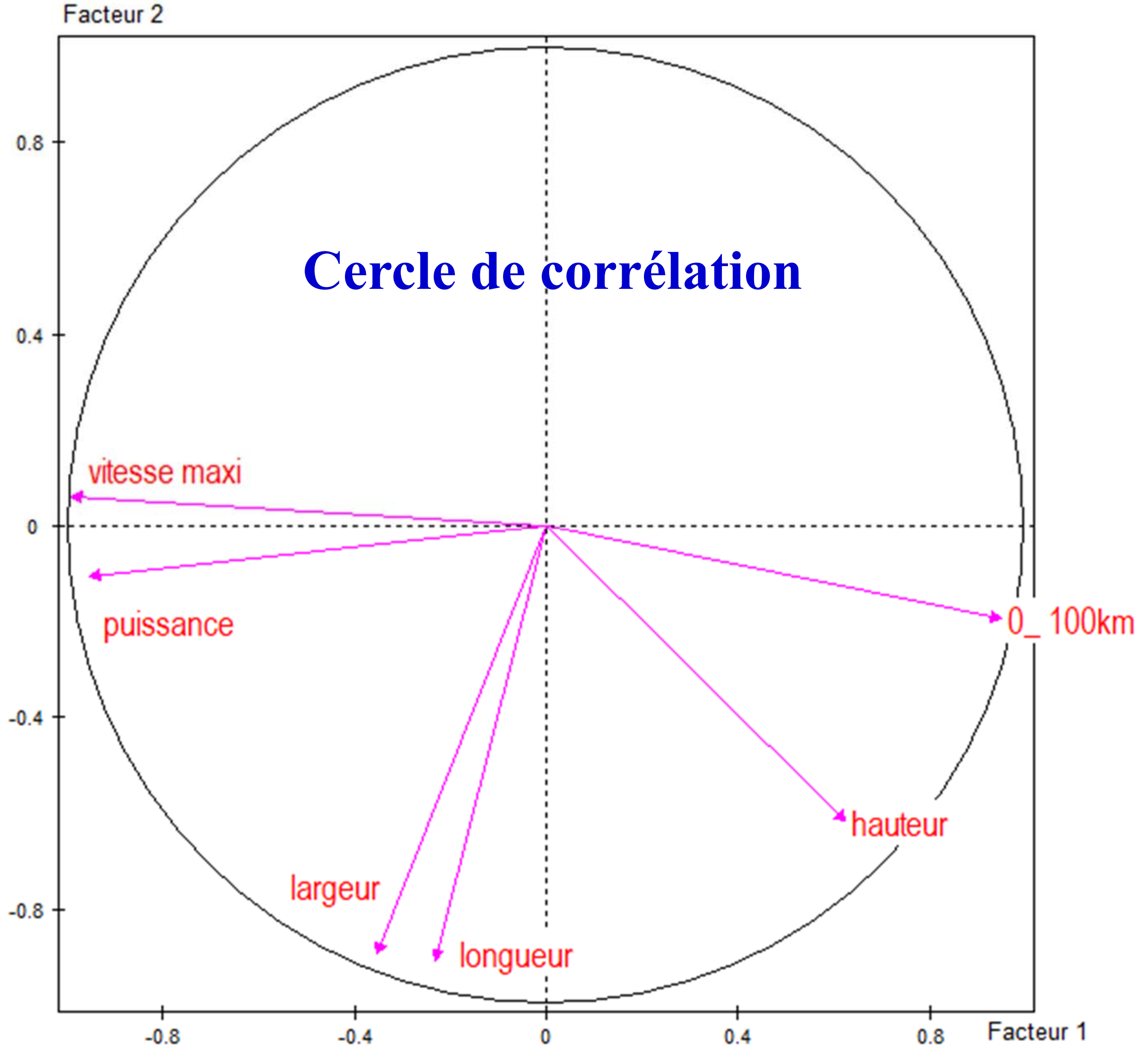
HISTOGRAMME DES 6 PREMIERES VALEURS PROPRES

NUMERO	VALEUR PROPRE	POURCENT.	POURCENT. CUMULE	
1	3.3569	55.95	55.95	**
2	2.0499	34.17	90.11	**
3	0.3612	6.02	96.13	*********
4	0.1365	2.28	98.41	****
5	0.0872	1.45	99.86	***
6	0.0083	0.14	100.00	*

COORDONNEES DES VARIABLES SUR LES AXES 1 A 5 VARIABLES ACTIVES

	COORDONNEES					CORRELATIONS VARIABLE-FACTEUR					ANCIENS AXES UNITAIRES				
LIBELLE COURT	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
C2 - longueur	-0.23	-0.91	-0.28	0.21	-0.03	-0.23	-0.91	-0.28	0.21	-0.03	-0.13	-0.63	-0.46	0.58	-0.12
C3 - largeur	-0.35	-0.89	-0.03	-0.28	-0.01	-0.35	-0.89	-0.03	-0.28	-0.01	-0.19	-0.62	-0.05	-0.75	-0.03
C4 - hauteur	0.62	-0.61	0.48	0.08	-0.04	0.62	-0.61	0.48	0.08	-0.04	0.34	-0.43	0.79	0.23	-0.13
C5 - puissance	-0.95	-0.11	0.18	0.04	0.21	-0.95	-0.11	0.18	0.04	0.21	-0.52	-0.07	0.30	0.10	0.71
C6 - vitesse maxi	-0.99	0.06	0.07	0.08	-0.02	-0.99	0.06	0.07	0.08	-0.02	-0.54	0.04	0.12	0.21	-0.08
C7 - 0_ 100km	0.95	-0.20	-0.14	0.01	0.20	0.95	-0.20	-0.14	0.01	0.20	0.52	-0.14	-0.23	0.03	0.68

Analyse en Composantes Principales avec SPAD



Analyse en Composantes Principales avec SPAD

Coheris SPAD

COORDONNEES, CONTRIBUTIONS ET COSINUS CARRES DES INDIVIDUS
AXES 1 A 5

INDIVIDUS			COORDONNEES					CONTRIBUTIONS					COSINUS CARRES				
IDENTIFICATEUR	P.REL	DISTO	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
01	5.00	2.02	1.12	0.00	-0.84	0.24	-0.08	1.9	0.0	9.7	2.2	0.3	0.62	0.00	0.35	0.03	0.00
02	5.00	14.30	2.85	2.31	-0.53	0.08	0.75	12.1	13.0	4.0	0.2	31.9	0.57	0.37	0.02	0.00	0.04
03	5.00	6.97	1.32	2.27	0.07	0.20	-0.14	2.6	12.5	0.1	1.4	1.2	0.25	0.74	0.00	0.01	0.00
04	5.00	10.11	2.20	2.18	0.66	-0.25	-0.07	7.2	11.6	6.0	2.2	0.3	0.48	0.47	0.04	0.01	0.00
05	5.00	5.97	2.20	-0.71	0.80	-0.06	0.01	7.2	1.2	8.8	0.1	0.0	0.81	0.08	0.11	0.00	0.00
06	5.00	0.97	0.88	0.10	0.23	-0.34	-0.14	1.2	0.0	0.7	4.3	1.2	0.80	0.01	0.05	0.12	0.02
07	5.00	1.52	-0.14	1.20	0.01	0.01	-0.26	0.0	3.5	0.0	0.0	3.9	0.01	0.94	0.00	0.00	0.04
08	5.00	0.63	0.49	0.50	-0.03	-0.31	-0.22	0.4	0.6	0.0	3.6	2.8	0.37	0.40	0.00	0.15	0.08
09	5.00	5.49	1.50	-1.54	0.72	0.57	0.03	3.4	5.8	7.2	11.9	0.1	0.41	0.43	0.10	0.06	0.00
10	5.00	0.62	0.01	-0.19	-0.71	0.18	-0.19	0.0	0.1	7.0	1.1	2.2	0.00	0.06	0.81	0.05	0.06
11	5.00	1.71	-1.13	0.56	0.00	-0.17	-0.29	1.9	0.8	0.0	1.1	4.9	0.74	0.18	0.00	0.02	0.05
12	5.00	1.49	0.69	-0.79	-0.62	0.01	0.02	0.7	1.5	5.4	0.0	0.0	0.32	0.42	0.26	0.00	0.00
13	5.00	16.45	1.76	-3.52	0.94	-0.33	0.10	4.6	30.1	12.3	4.1	0.6	0.19	0.75	0.05	0.01	0.00
14	5.00	3.87	-0.39	-1.64	-0.89	-0.47	-0.18	0.2	6.5	10.8	8.2	1.8	0.04	0.69	0.20	0.06	0.01
15	5.00	1.89	-1.31	-0.31	-0.19	0.11	-0.11	2.6	0.2	0.5	0.5	0.7	0.91	0.05	0.02	0.01	0.01
16	5.00	1.73	-0.47	-0.94	-0.53	0.44	-0.41	0.3	2.1	3.9	7.0	9.5	0.13	0.51	0.16	0.11	0.10
17	5.00	10.60	-3.06	0.37	0.48	0.92	0.06	14.0	0.3	3.2	31.1	0.2	0.88	0.01	0.02	0.08	0.00
18	5.00	12.15	-3.28	0.58	1.00	-0.01	0.27	16.0	0.8	13.7	0.0	4.1	0.88	0.03	0.08	0.00	0.01
19	5.00	15.54	-3.70	1.13	0.13	-0.76	0.07	20.4	3.1	0.2	20.9	0.3	0.88	0.08	0.00	0.04	0.00
20	5.00	5.94	-1.54	-1.57	-0.69	-0.05	0.77	3.5	6.0	6.5	0.1	34.2	0.40	0.42	0.08	0.00	0.10

Analyse en Composantes Principales avec SPAD

Projection des individus dans le premier plan factoriel