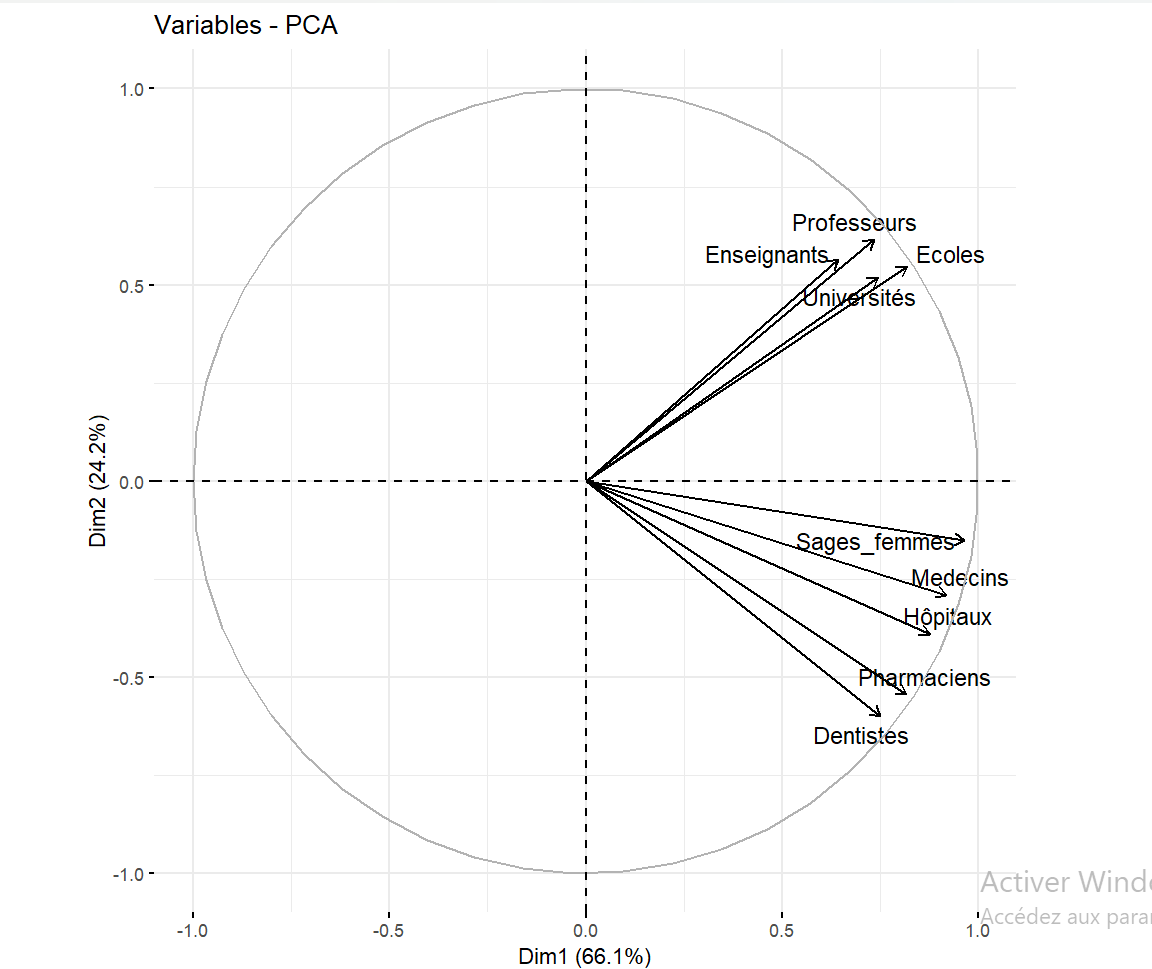
**EFFET TAILLE DANS UNE ACP**

On observe un effet de taille dans une ACP lorsque tous les points du nuage des variables sont regroupés du même côté d’un axe factoriel. L’exemple ci-dessous montre clairement un effet de taille car tous les points sont positivement corrélés à l’axe 1 du plan factoriel. Dans ce cas de figure, la première composante principale (ou premier axe factoriel) définit « *un facteur de taille* » et la deuxième composante principale différencie alors des individus de « taille » semblable : il définit un « *facteur de forme* ».



Un effet taille apparait généralement lorsque toutes les variables considérées ont des effectifs corrélés positivement entre elles. Supposons par exemple, effectuée une ACP sur des variables telles que : l’effectif de la population, le nombre de médecins, d’enseignements, d’élèves de chômeurs, … ces variables sont toutes corrélées positivement avec la démographie de la région. Et justement l’axe 1 sur lesquels il se regroupe n’est rien d’autre que la variable fictive « démographie ».

Voir <http://www.jybaudot.fr/Analdonnees/taille.html> pour approfondir l’exemple.

Une telle ACP ne permettra de distinguer ces variables que suivant leur ***taille*** (les flèches les plus courtes traduisant une démographie ou des quantités plus faibles et celles plus longues une plus forte démographie ou quantité). Et dans ce cas, selon que ces flèches se trouvent au-dessus ou en-dessous de l’axe 1, le second axe ne nous renseignera que sur la ***forme ou le type*** des variables considérées[[1]](#footnote-1).

Ainsi une ACP, dans laquelle un effet taille est détecté, est très pauvre en informations car ne nous donne pas plus que ça de spécificités sur notre échantillon mis à part la taille et la forme. C’est la raison pour laquelle il existe plusieurs méthodes qui ont été développées afin de « corriger » un effet taille. Les plus communes consistent à :

* Transformer les variables quantitatives en valeurs catégoriques et utiliser une autre méthode d’analyse factorielle, en l’occurrence l’analyse des correspondances multiples (ACM) ;
* Délaisser l’axe factoriel à l’origine de l’effet taille constaté ;
* Utiliser les pourcentages et non les données brutes ;
* Utiliser une ACP de type Varimax ;
* Faire une [ACP sur les rangs](http://www.jybaudot.fr/Analdonnees/acprangs.html) ;

<http://www.jybaudot.fr/Analdonnees/acprangs.html>

<http://www.statsoft.fr/concepts-statistiques/analyse-factorielle/analyse-factorielle.htm#strategies>

<https://sites.google.com/site/rgraphiques/4--stat/machine-learning-biostatistiques-analyse-de-donn%C3%A9es/analyse-en-composantes-principales/la-rotation-varimax>

1. Dans l’exemple, le second axe distingue les professions en liens avec la santé de ceux en lien avec l’éducation. [↑](#footnote-ref-1)