

INTERROGATION

Analyse Exploratoire des Données

Correction

Question 1 *Quel est le couple de variables qui sont les plus corrélées entre elles? Les moins corrélées entre elles? Quelles sont les variables les plus opposées?*

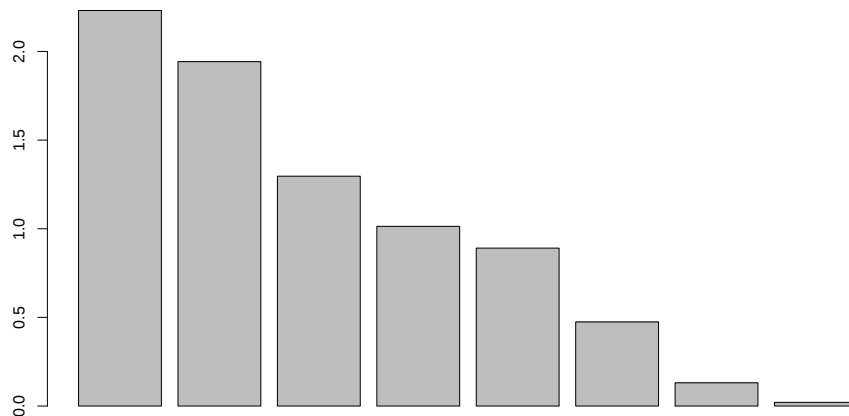
Le couple le plus corrélé ($r \approx 1$) est (**SensiDCS**, **SensiGSM**) avec 0,93.

Le couple le moins corrélé ($r \approx 0$) est (**AutoConv**, **Commodité**) avec $-0,01$.

Le couple le plus opposé ($r \approx -1$) est (**PuisGSM**, **Acoustiq**) avec $-0,43$.

Question 2 *Faire une représentation en histogramme des valeurs propres. Combien de composantes principales faut-il retenir? Quel est le pourcentage d'inertie totale expliquée par le sous-espace principal correspondant? Si on ne retient que le premier plan principal, quel est le ratio expliqué?*

On obtient l'histogramme suivant pour les valeurs propres :



Le critère de Kaiser ($\lambda > 1$) amène à considérer les 3 premières valeurs propres (on pourrait en conserver 4, mais la quatrième valeur est vraiment proche de 1). L'inertie totale est 8 et l'inertie portée par les 3 premiers axes est $2,23 + 1,94 + 1,30 = 5,47$. Cela représente 68% de l'inertie totale.

Si on ne considère que les deux premières valeurs propres, on n'a que 52% de l'inertie totale.

Question 3 *Quelles sont les variables qui déterminent la première composante principale? La proximité des ces variables sur le cercle est-elle cohérente avec ce que l'on sait des données d'origine?*

La première composante principale est déterminée par **SensiGSM** et **SensiDCS**. On le voit sur le cercle, mais aussi par les corrélations, respectivement $-0,96$ et $-0,95$.

Comme ces variables sont très près du cercle, on peut déduire de leur proximité dans le cercle de corrélation qu'elles sont très corrélées l'une à l'autre. En effet, ce sont les deux variables notées à la question 1 comme ayant une corrélation de 0,93.

Question 4 *Quels sont les individus qui déterminent le plus la première composante principale ? Quelle est leur contribution ?*

Les individus qui déterminent la première composante principale sont **Panasonic** et **Sagem**. On le voit sur la projection des individus mais surtout par leur contribution aux axes (qui sont données ici en 10000ièmes) :

- **Panasonic**: 0,3127
- **Sagem**: 0,3215

Ces valeurs sont à comparer avec le poids de chaque individu, qui est ici $1/15 \approx 0,07$. Elles sont très au dessus des valeurs normales, puisque les contributions des individus somment à 1 et que ces deux éléments à eux seuls en représentent près des deux tiers.

Question 5 *Quel est le lien entre les variables de la question 3 et les individus de la question 4 ?*

D'une part le premier axe est déterminé par **SensiGSM** et **SensiDCS** (question 3). D'autre part il est déterminé par les deux individus **Panasonic** et **Sagem** (question 4).

Il est en fait facile de vérifier que ces deux portables ont une note hors norme pour ces deux variables, ce qui explique les caractéristiques de la première composante principale.