

Analyse en Composantes Principales

Jeu de données scooter

Ce jeu de données contient 416 individus et 25 variables, 1 variable quantitative est illustrative, 6 variables qualitatives sont illustratives.

1. Observation d'individus extrêmes

L'analyse des graphes ne révèle aucun individu singulier.

2. Distribution de l'inertie

L'inertie des axes factoriels indique d'une part si les variables sont structurées et suggère d'autre part le nombre judicieux de composantes principales à étudier.

Les 2 premiers axes de l'analyse expriment **37.04%** de l'inertie totale du jeu de données ; cela signifie que 37.04% de la variabilité totale du nuage des individus (ou des variables) est représentée dans ce plan. C'est un pourcentage relativement moyen, et le premier plan représente donc seulement une part de la variabilité contenue dans l'ensemble du jeu de données actif. Cette valeur est nettement supérieure à la valeur référence de **15.52%**, la variabilité expliquée par ce plan est donc hautement significative (cette inertie de référence est le quantile 0.95-quantile de la distribution des pourcentages d'inertie obtenue en simulant 2523 jeux de données aléatoires de dimensions comparables sur la base d'une distribution normale).

Du fait de ces observations, il serait alors probablement nécessaire de considérer également les dimensions supérieures ou égales à la troisième dans l'analyse.

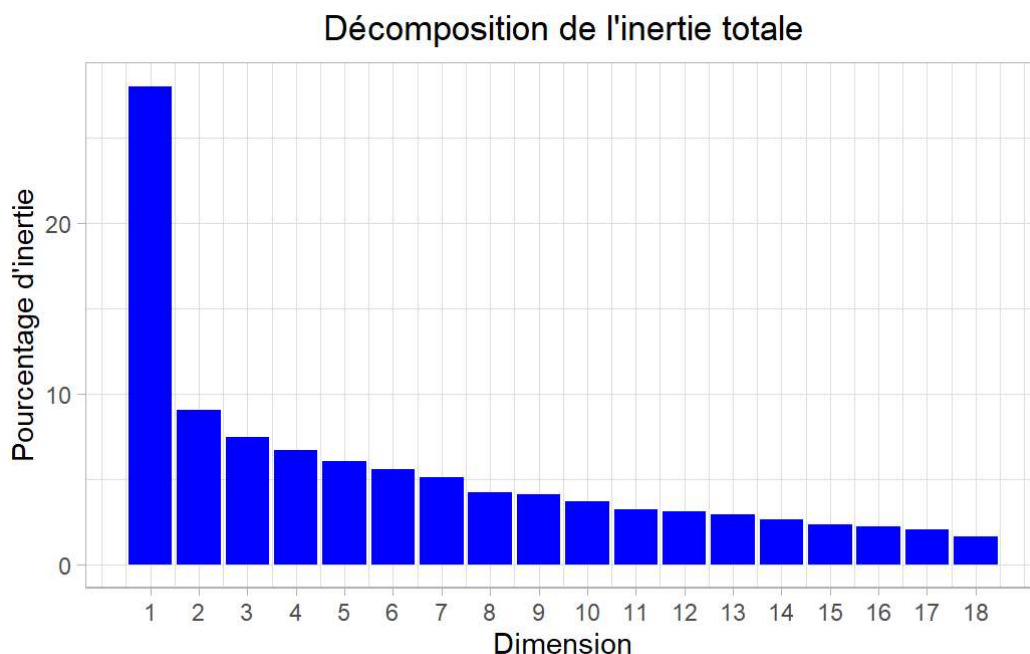


Figure 2 - Décomposition de l'inertie totale

Une estimation du nombre pertinent d'axes à interpréter suggère de restreindre l'analyse à la description des 3 premiers axes. Ces composantes révèlent un taux d'inertie supérieur à celle du quantile 0.95-quantile de distributions aléatoires (44.47% contre 22.58%). Cette observation suggère que seuls ces axes sont porteurs d'une véritable information. En conséquence, la description de l'analyse sera restreinte à ces seuls axes.

3. Description du plan 1:2

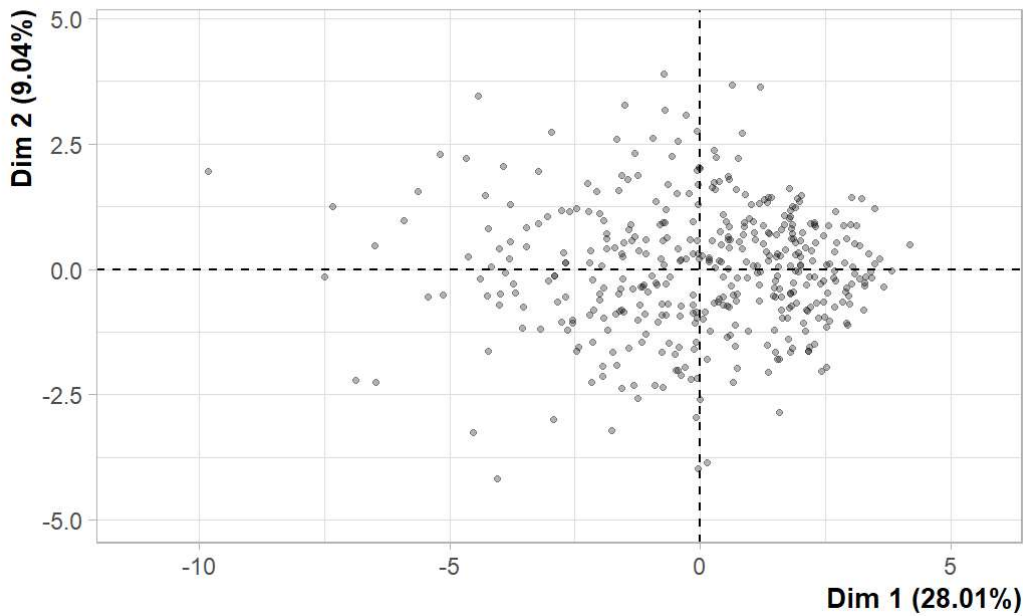


Figure 3.1 - Graphe des individus (ACP) Les individus libellés sont ceux ayant la plus grande contribution à la construction du plan.

La probabilité critique du test de Wilks indique la variable dont les modalités sépare au mieux les individus sur le plan (i.e. qui explique au mieux les distances entre individus).

##	type-cylindrée	sexe	critère-esthétique	CSP	
âge	type-utilisation				
##	5.908269e-17	7.623612e-07	2.047886e-02	1.999731e-01	2.761817
e-01	7.495597e-01				

La meilleure variable qualitative pour illustrer les distances entre individus sur le plan est la variable : *type-cylindrée*.

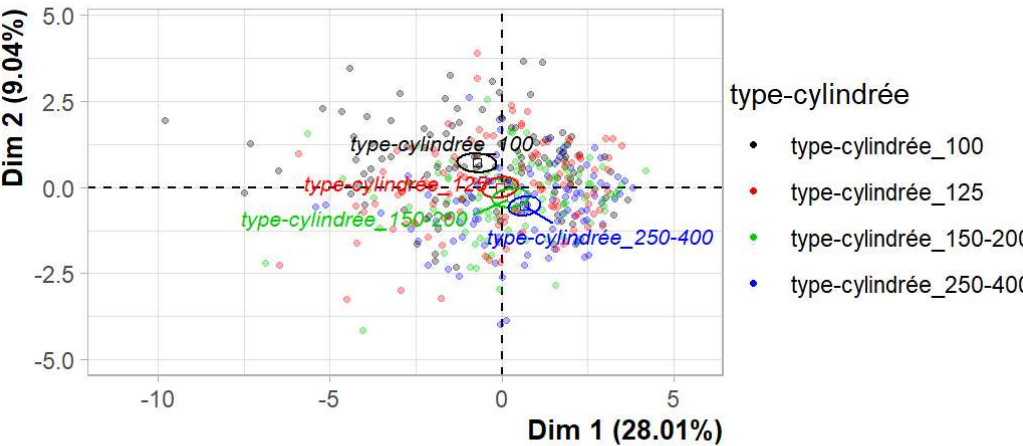


Figure 3.2 - Graphe des individus (ACP) Les individus libellés sont ceux ayant la plus grande contribution à la construction du plan. Les individus sont colorés selon leur appartenance aux modalités de la variable type-cylindrée.

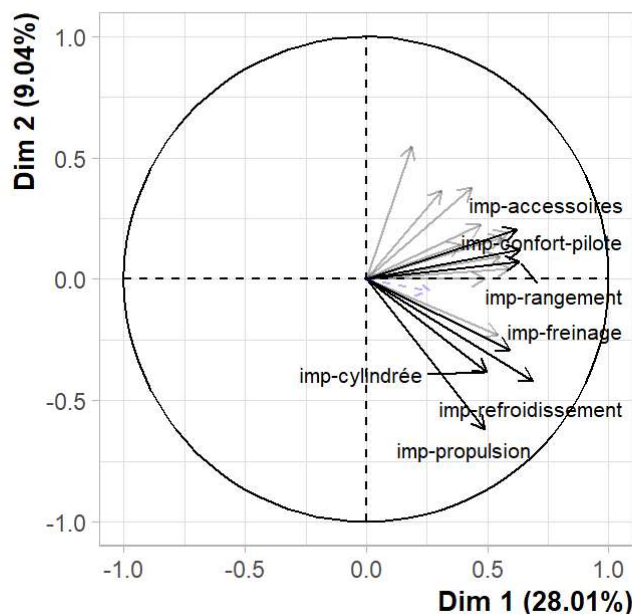


Figure 3.3 - Graphe des variables (ACP) Les variables en noir sont les variables actives, celles en bleu sont les variables quantitatives supplémentaires (illustratives). Les variables libellées sont celles les mieux représentées sur le plan.

```
## Warning: ggrepel: 5 unlabeled data points (too many overlaps). Consider increasing max.overlaps
```

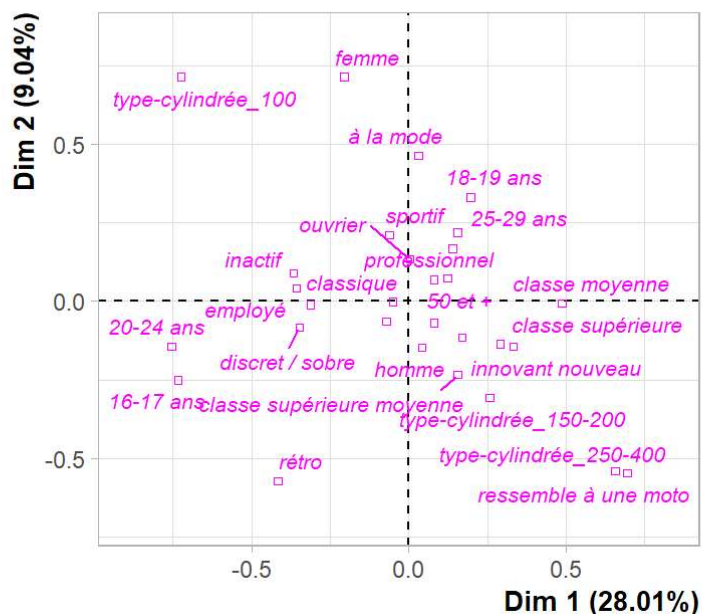


Figure 3.4 - Graphe des modalités (ACP) Les facteurs libellés sont ceux les mieux représentés sur le plan.

La **dimension 1** oppose des individus caractérisés par une coordonnée fortement positive sur l'axe (à droite du graphe) à des individus caractérisés par une coordonnée fortement négative sur l'axe (à gauche du graphe).

Le groupe 1 (caractérisés par une coordonnée positive sur l'axe) partage :

- de fortes valeurs pour des variables telles que *imp-refroidissement*, *imp-freinage*, *imp-confort-passager*, *imp-accessoires*, *imp-propulsion*, *imp-rangement*, *imp-tableau-de-bord*, *imp-fiabilité-moteur*, *imp-cylindrée* et *imp-confort-pilote* (de la plus extrême à la moins extrême).

Le groupe 2 (caractérisés par une coordonnées négative sur l'axe) partage :

- de faibles valeurs pour des variables telles que *imp-refroidissement*, *imp-accessoires*, *imp-confort-pilote*, *imp-freinage*, *imp-rangement*, *imp-confort-passager*, *imp-fiabilité-moteur*, *imp-tableau-de-bord*, *imp-cylindrée* et *imp-propulsion* (de la plus extrême à la moins extrême).

Le groupe 3 (caractérisés par une coordonnées négative sur l'axe) partage :

- de faibles valeurs pour des variables telles que *imp-prix*, *imp-antivol*, *imp-magasin*, *imp-tableau-de-bord*, *imp-confort-passager*, *imp-accessoires*, *imp-tablier-avant*, *imp-rangement*, *imp-feux* et *imp-esthétique* (de la plus extrême à la moins extrême).

La **dimension 2** oppose des individus caractérisés par une coordonnée fortement positive sur l'axe (en haut du graphe) à des individus caractérisés par une coordonnée fortement négative sur l'axe (en bas du graphe).

Le groupe 1 (caractérisés par une coordonnée positive sur l'axe) partage :

- de faibles valeurs pour des variables telles que *imp-refroidissement*, *imp-accessoires*, *imp-confort-pilote*, *imp-freinage*, *imp-rangement*, *imp-confort-passager*, *imp-fiabilité-moteur*, *imp-tableau-de-bord*, *imp-cylindrée* et *imp-propulsion* (de la plus extrême à la moins extrême).

Le groupe 2 (caractérisés par une coordonnée positive sur l'axe) partage :

- de fortes valeurs pour les variables *imp-magasin*, *imp-esthétique*, *imp-accessoires* et *imp-tableau-de-bord* (de la plus extrême à la moins extrême).
- de faibles valeurs pour les variables *imp-propulsion*, *imp-refroidissement*, *imp-freinage* et *imp-cylindrée* (de la plus extrême à la moins extrême).

Le groupe 3 (caractérisés par une coordonnées négative sur l'axe) partage :

- de faibles valeurs pour des variables telles que *imp-prix*, *imp-antivol*, *imp-magasin*, *imp-tableau-de-bord*, *imp-confort-passager*, *imp-accessoires*, *imp-tablier-avant*, *imp-rangement*, *imp-feux* et *imp-esthétique* (de la plus extrême à la moins extrême).

Le groupe 4 (caractérisés par une coordonnées négative sur l'axe) partage :

- de fortes valeurs pour des variables telles que *imp-refroidissement*, *imp-freinage*, *imp-confort-passager*, *imp-accessoires*, *imp-propulsion*, *imp-rangement*, *imp-tableau-de-bord*, *imp-fiabilité-moteur*, *imp-cylindrée* et *imp-confort-pilote* (de la plus extrême à la moins extrême).
-

4. Description de la dimension 3

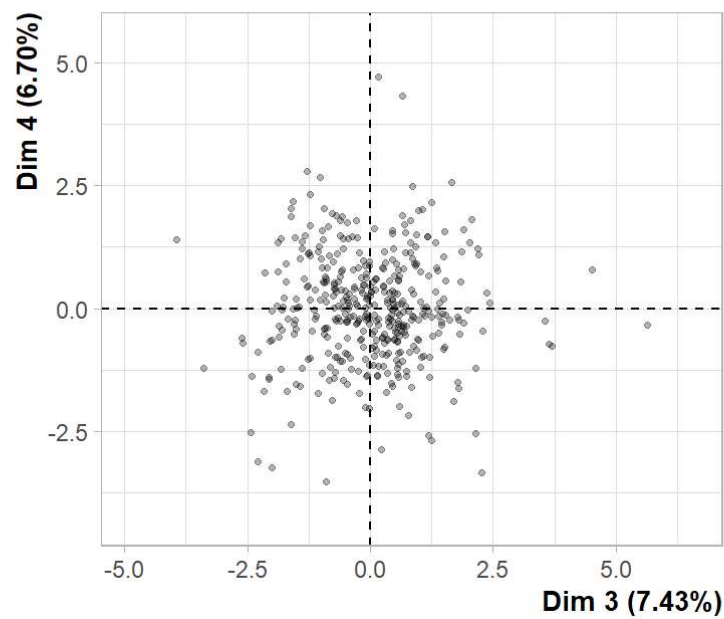


Figure 4.1 - Graphe des individus (ACP) Les individus libellés sont ceux ayant la plus grande contribution à la construction du plan.

La probabilité critique du test de Wilks indique la variable dont les modalités sépare au mieux les individus sur le plan (i.e. qui explique au mieux les distances entre individus).

##	critère-esthétique	type-utilisation	type-cylindrée	âge	
CSP	sexe				
##	6.791642e-06	6.635396e-03	1.151155e-02	2.062246e-01	4.639292
e-01	6.783833e-01				

La meilleure variable qualitative pour illustrer les distances entre individus sur le plan est la variable : *critère-esthétique*.

```
## Warning: ggrepel: 6 unlabeled data points (too many overlaps). Consider increasing max.overlaps
```

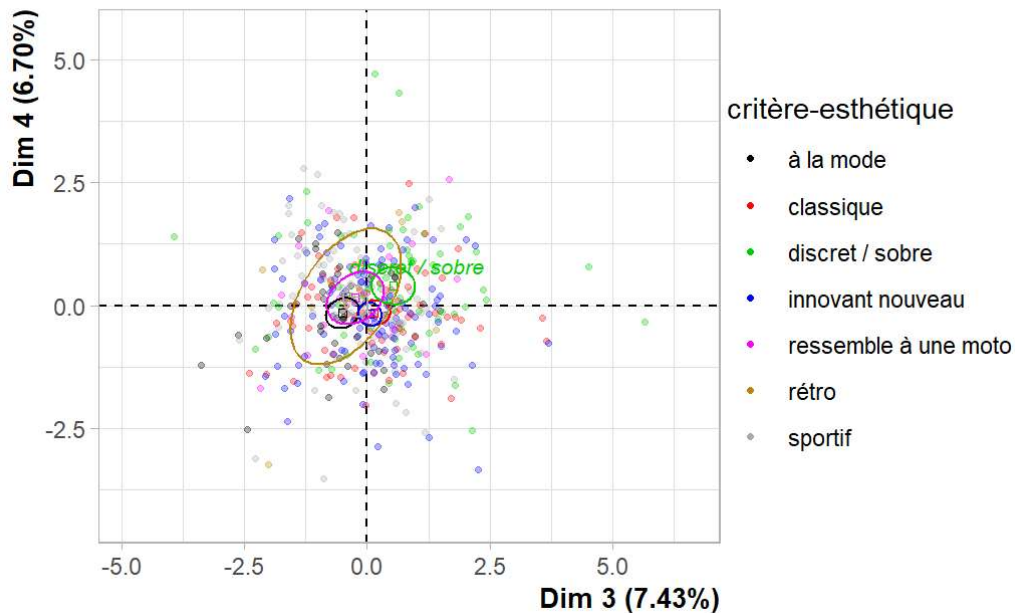


Figure 4.2 - Graphe des individus (ACP) Les individus libellés sont ceux ayant la plus grande contribution à la construction du plan. Les individus sont colorés selon leur appartenance aux modalités de la variable critère-esthétique.

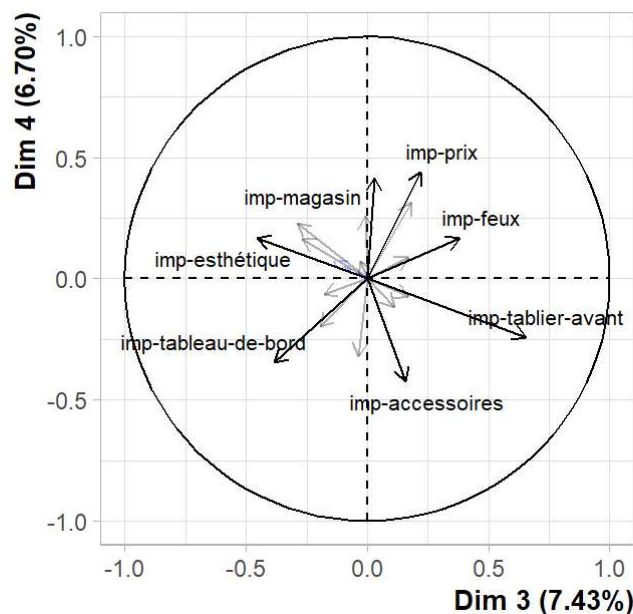


Figure 4.3 - Graphe des variables (ACP) Les variables en noir sont les variables actives, celles en bleu sont les variables quantitatives supplémentaires (illustratives). Les variables libellées sont celles les mieux représentées sur le plan.

```
## Warning: ggrepel: 9 unlabeled data points (too many overlaps). Consider increasing max.overlaps
```

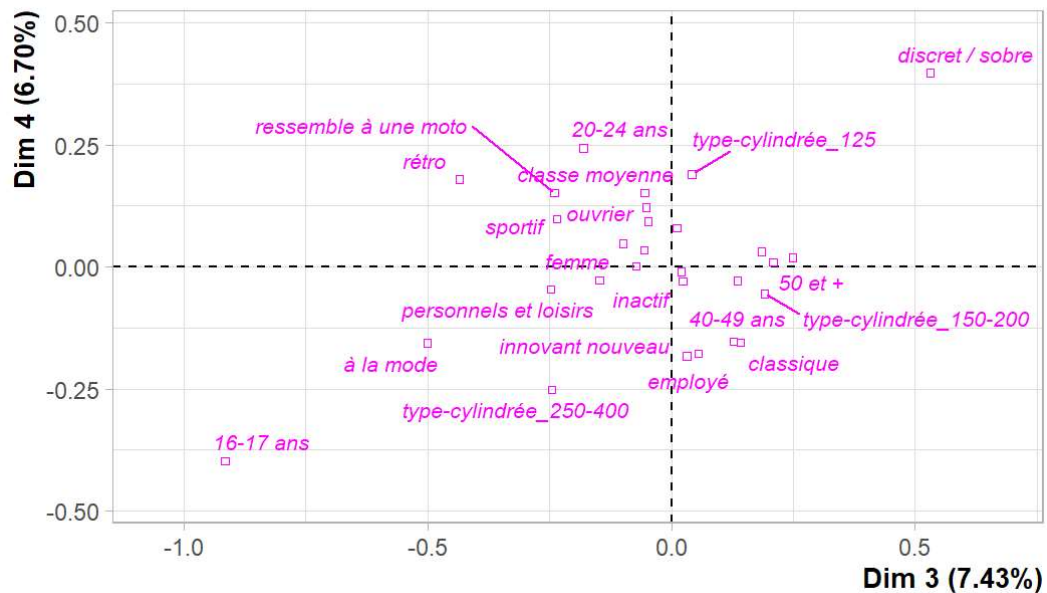


Figure 4.4 - Graphe des modalités (ACP) *Les facteurs libellés sont ceux les mieux représentés sur le plan.*

La **dimension 3** oppose des individus caractérisés par une coordonnée fortement positive sur l'axe (à droite du graphe) à des individus caractérisés par une coordonnée fortement négative sur l'axe (à gauche du graphe).

Le groupe 1 (caractérisés par une coordonnée positive sur l'axe) partage :

- de fortes valeurs pour les variables *imp-tablier-avant*, *imp-accessoires*, *imp-antivol*, *imp-feux*, *imp-rangement* et *imp-refroidissement* (de la plus extrême à la moins extrême).
- de faibles valeurs pour les variables *imp-esthétique*, *imp-marque*, *imp-cylindrée* et *imp-magasin* (de la plus extrême à la moins extrême).

Le groupe 2 (caractérisés par une coordonnées négative sur l'axe) partage :

- de fortes valeurs pour la variable *imp-tableau-de-bord*.
- de faibles valeurs pour les variables *imp-prix*, *imp-feux*, *imp-magasin*, *imp-fiabilité-moteur*, *imp-dimensions* et *imp-tablier-avant* (de la plus extrême à la moins extrême).

Le groupe 3 (caractérisés par une coordonnées négative sur l'axe) partage :

- de fortes valeurs pour les variables *imp-magasin*, *imp-prix*, *imp-dimensions*, *imp-esthétique*, *imp-marque*, *imp-fiabilité-moteur*, *imp-cylindrée* et *imp-feux* (de la plus extrême à la moins extrême).
- de faibles valeurs pour les variables *imp-accessoires*, *imp-tablier-avant*, *imp-antivol* et *imp-tableau-de-bord* (de la plus extrême à la moins extrême).

Annexes