

Exercice 1 Le but est d'étudier la répartition du temps passé sur différentes activités d'après les données de la table onu67 en faisant une ACP avec R. Le script que vous allez rédiger pour cet exercice pourra vous servir de base pour réaliser d'autres analyses.

Chargement et première description des données

Question 1.1 Charger la table sous R. Calculer un tableau T_s des variables numériques centrées et **réduites** (on fait donc une ACP normée ; la fonction scale pourra être utile).

Question 1.2 Calculer l'inertie de chaque individu, et de l'ensemble du nuage – par rapport au centre d'inertie (fonctions sum, rowSums par exemple).

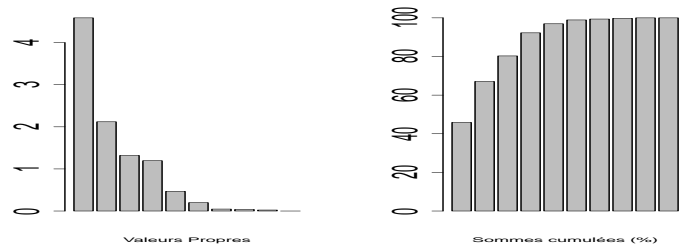
Question 1.3 Calculer la matrice des corrélations entre variables (fonction cor ou cov). Quelles variables semblent corrélées ?

Calcul des axes factoriels

Question 1.4 Diagonaliser la matrice des corrélations avec la fonction eigen : si on appelle E le résultat de la fonction eigen, alors Evalues$ et Evector$ s contiennent respectivement les valeurs propres et les vecteurs propres de la matrice diagonalisée.

Les valeurs propres correspondent aux inerties des axes factoriels. Il n'y a pas de règle absolue concernant le nombre d'axes à retenir. Une règle courante est de retenir suffisamment d'axes pour avoir 80% de l'inertie. Une autre règle est d'éliminer chaque axe dont l'inertie est inférieure à la moyenne. Une méthode consiste à visualiser les valeurs propres par ordre décroissant, et à chercher un point où la « courbe » semble indiquer un saut.

Question 1.5 Afin d'avoir une idée du nombre d'axes factoriels à considérer dans la suite, réaliser un diagramme présentant les valeurs propres (fonction barplot) ainsi qu'un diagramme présentant les pourcentages cumulés de ces valeurs propres (fonction cumsum notamment).



Combien d'axes vous semblent nécessaires ici ?

Projection des individus

Les vecteurs propres forment un nouveau repère, on veut projeter le nuage d'individus sur les plans déterminés par certains de ces vecteurs.

Question 1.6 Calculer le tableau T_p des coordonnées des individus dans le nouveau repère.

Question 1.7 Afficher les projections des individus sur le plan correspondant aux deux premiers axes factoriels.

Question 1.8 Ajouter des points représentant des « individus moyens », par exemple l'homme moyen et la femme moyenne.

Représentation des variables

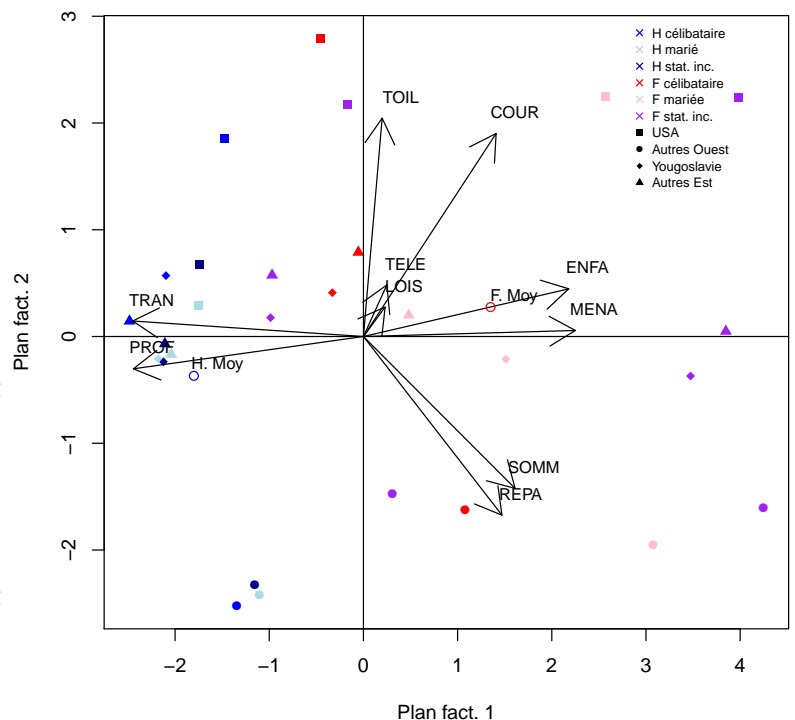
Question 1.9 Calculer les corrélations entre les variables initiales et les variables correspondant aux axes factoriels.

Question 1.10 Ajouter sur le graphique précédent les variables initiales, représentées par des flèches depuis l'origine, avec leurs noms.

Question 1.11 Encapsuler le code R des questions 7 à 10 dans une fonction, de manière à réaliser facilement des projections sur d'autres plans factoriels.

Question 1.12 Que peut-on conclure sur les temps d'activités de toutes ces catégories de personnes en 1967 ?

Enquête ONU 1967



Qualités de représentation

La qualité de la représentation H d'un individu I sur un axe est définie comme étant le rapport $\frac{OH^2}{OI^2}$ – c'est le carré du cosinus de l'angle entre les segments OI et OH : plus ce cosinus est proche de 1, plus l'angle est faible, et donc plus I est proche de sa projection H . La qualité de représentation de I sur un plan défini par deux axes de projection est la somme des qualités de représentation sur chacun des deux axes.

Question 1.13 Calculer les qualités de représentation des individus sur l'axe 1, puis sur le plan contenant les axes 1 et 2.

Étude des axes

Question 1.14 Calculer un vecteur donnant l'inertie de chaque individu sur l'axe 1. Calculer un vecteur donnant pour chaque individu sa *contribution* à la formation de cet axe 1 : c'est le pourcentage de l'inertie de l'individu sur l'axe 1, par rapport à la somme des inerties de tous les individus sur cet axe.

Question 1.15 Calculer les contributions des variables à la formation de l'axe 1 : pour chaque variable initiale, c'est le pourcentage du carré de sa corrélation avec l'axe 1 par rapport à la somme des carrés des corrélations de toutes les variables avec l'axe 1.

Question 1.16 Interpréter l'axe 1 suivant les individus et variables qui contribuent le plus à sa formation : quelles variables oppose-t-il ? Quels individus oppose-t-il ?

Question 1.17 Interpréter de même les autres axes conservés (cf. question 1.5).