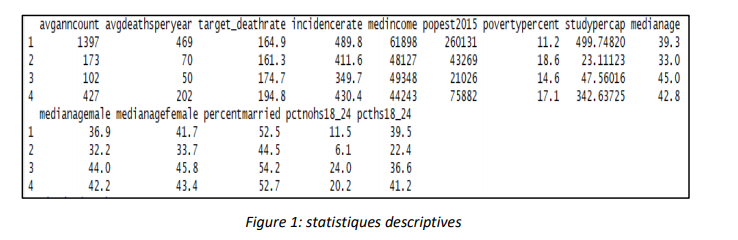
ACP



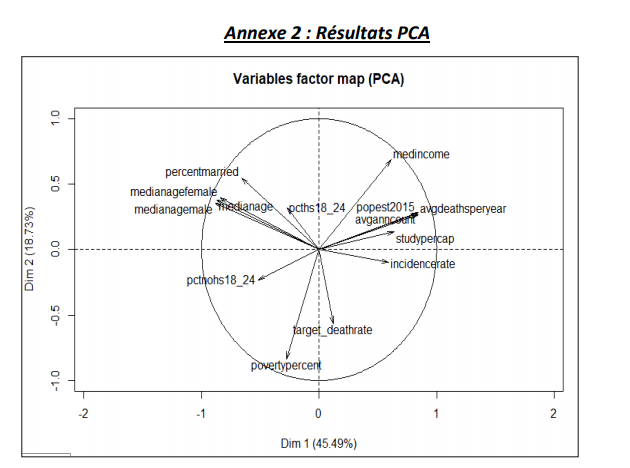


Figure2 : cercle de corrélation

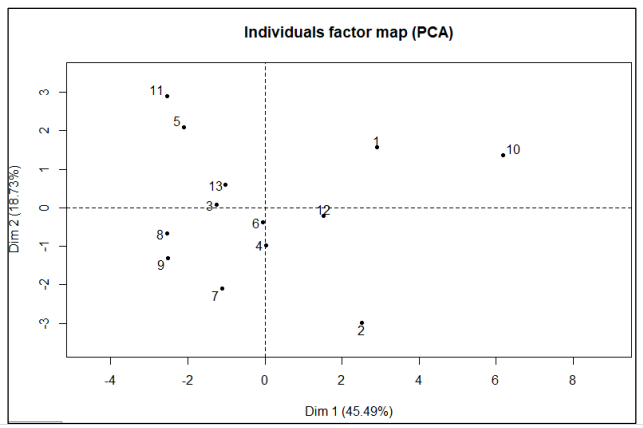
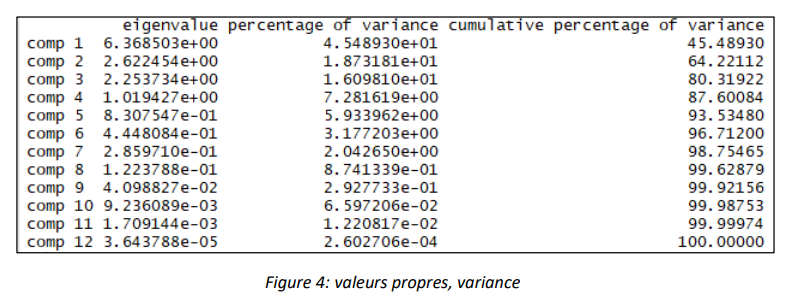
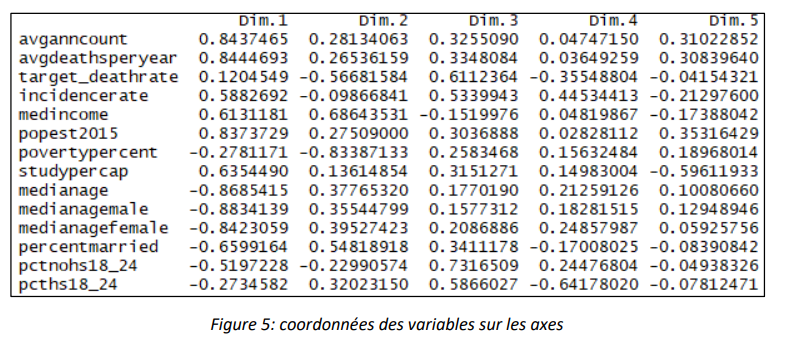
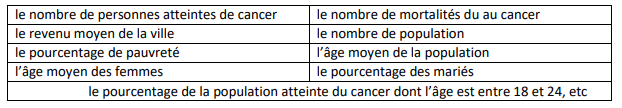


Figure 3 : carte des individus





On donne sur l’annexe 1 un échantillon de données pour quelques villes. Ces dernières sont décrites par :



Les statistiques descriptives sont données par la figure 1. On se propose de réaliser une analyse en composantes principales afin de comprendre les données. Les résultats sont illustrés dans les figures 2,3,4 et5.

1. Définir l’analyse en composantes principales et préciser son utilité.

L’analyse en composante principale est une méthode d’analyse descriptive qui permet de réduire la dimensionnalité de nos données afin de pouvoir les projeter sur un plan factoriel. Cette projection est très utile car elle permet de souligner la corrélation qui existe entre variable/ variable, individu/ individu, axe principal/ variable ou axe principal/individu. Des interprétations très importantes peuvent avoir lieu suite à ces projections, par exemple la détection de la variable pertinente qui caractérise un ensemble d’individus. Par ailleurs, elle peut nous mener à établir une segmentation dans certains cas. Elle est aussi utilisée pour synthétiser les variables d’origines (grand nombre de variables) puisqu’elle peut éliminer toute redondance au niveau des données pour aboutir à des nouvelles variables qui sont une combinaison linéaire des variables originaux (les axes principaux sont orthogonaux).

1. En se référant au tableau des valeurs propres donné par la figure 4, comment choisir les axes factoriels les plus adéquats. Quel est le critère utilisé ?

Le tableau visualise les valeurs propres et selon le critère de kaiser il faut choisir les valeurs propres qui sont supérieures à 1. Il s’agit alors de choisir les quatre premières composantes principales.

1. Que représente la variance ? comment la calculer ?

La variance expliquée par une composante principale correspond à l’information portée par cet axe et elle est calculée comme suit : valeur propre de la composante principale/ total des valeurs propres.

1. Pour des raisons de visualisation, on a choisi de représenter nos variables sur les axes Dim1 et Dim2. Comment jugez-vous ce repère de projection. Interpréter les corrélations variables/ variables et variables /dimensions.

La variance cumulée et expliquée par les deux axes est égale à 64.22. Donc ce n’est pas un repère assez fiable car il cache plusieurs informations sur les données.

1. Quelles sont le(s) variable(s) qui contribuent fortement à la formation de la première composante principale (selon le tableau 5)

Pour ce faire il faut calculer la moyenne des projections des variables sur la première dimension et puis garder ceux qui sont supérieur à la moyenne : la moyenne est 0.62 donc les variables qui contribuent fortement à la formation de la 1ère CP est :

+ : avganncount, avgdeathsperyear, popest 2015, studypercap

- :medianage, medianagemale, medianagefemale, percentmarried

1. Confirmer la réponse précédente en vous référant à la figure 2.

D’après la figure, les variables les plus corrélées avec la première CP sont les mêmes qu’énoncé précédemment car soient elles forment un angle très étroit avec cette CP (corrélation positive) ou un angle très obtus (corrélation négative).

1. Quels sont les individus qui sont bien représentés dans le plan factoriel.

11, 10 et 2.