

Master 1 : Économétrie et statistiques
Analyse de données et descriptive

CLASSIFICATION DE 63 PAYS EN
DEVELOPPEMENT SUR L'ANNEE 2020
SUIVANT LE TAUX DE MORTALITE
INFANTO-JUVENILE

I. NOTRE BASE DE DONNEES

Notre projet réalisé dans le cadre du cours d'analyse de données portait sur l'étude du taux de mortalité infanto-juvénile dans 63 pays en développement sur l'année 2020.

Notre travail a consisté à réaliser une analyse en composantes principales afin d'identifier les facteurs justifiant la disparité du taux de mortalité des enfants âgés de moins de 5 ans au sein des pays en développement.

Afin d'avoir une étude sur les données les plus récentes possibles, nous avons utilisé les données fournies par la banque mondiale et l'OMS (un résumé statistique est en annexe). Notre jeu de données est composé de 63 pays et de 13 variables quantitatives suivantes: taux de mortalité des enfants de – de 5 ans (en%), PIB/habitant, taux de la maigreur de l'enfant (en%), prévalence de l'hépatite B (appelé le VHB) chez l'enfant, couverture vaccinale diphtérie-tétanos-coqueluche (DTC3) chez les enfants ayant 1 an (en%), prévalence de l'anémie chez les femmes en âge de procréer (15 à 49 ans) en %, taux de scolarisation des filles dans le primaire, densité de médecins pour 10 000 habitants (en%), densité d'infirmiers et de personnels obstétriques pour 10 000 habitants (en%), dépenses publiques pour la santé (en%), total net de l'aide publique au développement destinée aux secteurs de la recherche médicale et de la santé de base, par habitant (en dollars américains), par pays bénéficiaire, accès aux services d'eau de base (en% de la population) ainsi que l'espérance de vie à la naissance.

II. CLASSIFICATION

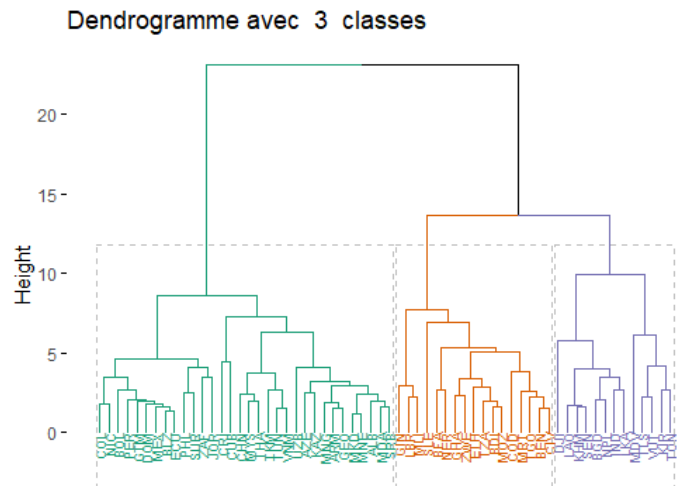
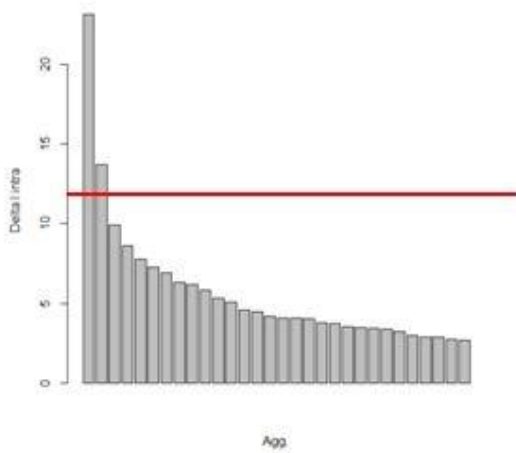
La classification consiste à construire des classes ou groupes. Une classe est un sous-ensemble d'objets ou d'individus (ici : pays) qui se ressemblent tout en se différenciant des autres pays placés dans les autres classes. Cette procédure va nous permettre de savoir si on peut dégager des profils types de pays selon la mortalité infanto-juvénile.

a) Classification ascendante hiérarchique (CAH)

Cette méthode de classification permet de déterminer le nombre de classes avec le critère d'agrégation (Ward) et de construire un dendrogramme mettant en évidence des liens hiérarchiques entre les pays. La classification est de bonne qualité si et seulement si les conditions suivantes sont vérifiées :

- Homogénéité des classes
- Les classes sont séparées
- Le nombre de classes est correct

A la vue de l'histogramme, nous pouvons voir qu'il y a un écart important de la variation de l'inertie intra-classes. Ce dernier est synonyme d'une forte inertie intra-classes et d'une faible inertie inter-classes. En d'autres termes, les principes d'homogénéité et de séparation du critère Ward ne sont pas vérifiés. Par conséquent, nous avons décidé de s'arrêter avant cet écart et avons obtenu un nombre optimal de 3 classes grâce à l'affichage du dendrogramme et à la confirmation du logiciel R avec l'utilisation de la commande nbClust. La qualité de la partition est de 46%.

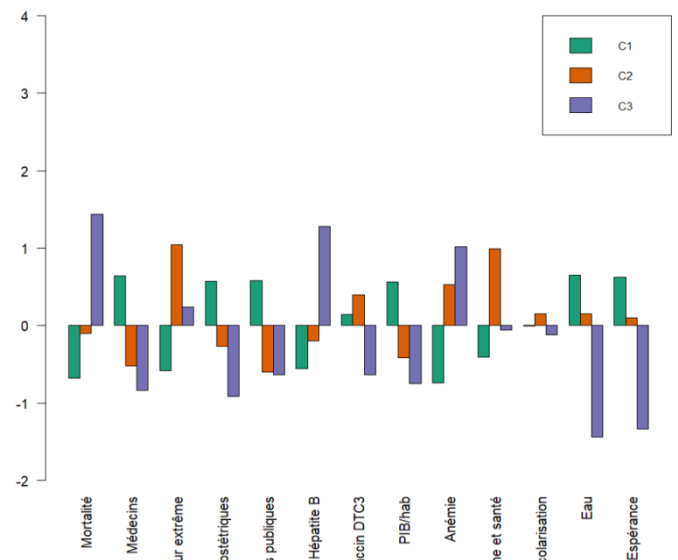
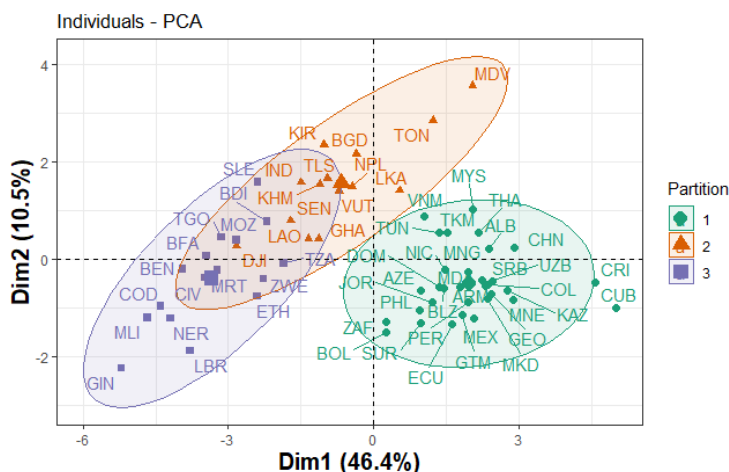


b) Méthode de partitionnement (K-means)

La partition obtenue par CAH n'est pas toujours optimale et peut être améliorée par les K-means. Cette méthode possède un algorithme itératif en prenant la partition de CAH comme initialisation et se présente comme suit :

- Etape de représentation = calcul des barycentres des classes
- Etape d'affectation = réaffecter les individus à la classe dont ils sont le plus proches du barycentre.

L'algorithme s'arrête lorsqu'il y a convergence.



Dans notre études, la qualité de la partition reste égale a 46%. Nous observons que sur le plan factoriel, l'axe 1 discrimine C1 et l'axe 2 discrimine C2 et C3. Enfin, l'étude des barycentres nous a permis de définir 3 groupes de pays de la façon suivante :

Classe des Pays	Pays le plus proche du barycentre	Nombre de pays	Variables positives	Variables négatives
C1= très bonne situation sanitaire	MKD = Macédoine du Nord	32	Médecins, Infirmiers, Eau, Espérance de vie	Mortalité, Anémie, Maigreur
C2 = assez bonne situation sanitaire	TLS = Timor-Leste	14	Maigreur, Anémie, Recherche	Mortalité, Médecins, Dépenses publiques
C3 = mauvaise situation sanitaire	CIV = Côte d'Ivoire	16	Mortalité, Hépatite B, Anémie	Médecins, Eau, Espérance de vie, Infirmiers

Les pays des différentes classes ont des similitudes avec le pays qui est le plus proche du barycentre. Ainsi, les pays les plus sensibles d'avoir un taux élevé de mortalité ont des caractéristiques communes avec la Côte d'Ivoire.

III. ANNEXE

Tableau des statistique des variables :

Characteristic	N = 62 ⁱ
Mortalité	Moy = 34 [Etendue = 108-2] écart-type = 28 Q1 = 12 médiane=25 et Q3= 50
Médecins	Moy = 14 [Etendue = 84-0] écart-type = 16 Q1 = 2 médiane=9 et Q3= 23
Maigreux extrême	Moy = 5.4 [Etendue = 21.5-0.4] écart-type = 4.2 Q1 = 2.1 médiane=4.3 et Q3= 8.0
Infirmiers et personnels obstétriques	Moy = 28 [Etendue = 113-2] écart-type = 22 Q1 = 10 médiane=24 et Q3= 41
Dépenses publiques	Moy = 9.8 [Etendue = 25.2-3.1] écart-type = 4.5 Q1 = 6.5 médiane=8.6 et Q3= 12.6
Prévalence Hépatite B	Moy = 1.06 [Etendue = 6.07-0.02] écart-type = 1.41 Q1 = 0.11 médiane=0.36 et Q3= 1.59
Vaccin DTC3	Moy = 84 [Etendue = 99-47] écart-type = 12 Q1 = 78 médiane=86 et Q3= 93
PIB/hab	Moy = 3,974 [Etendue = 12,179-217] écart-type = 3,055 Q1 = 1,492 médiane=3,361 et Q3= 5,570
Anémie	Moy = 32 [Etendue = 59-7] écart-type = 13 Q1 = 21 médiane=29 et Q3= 42
Recherche et santé	Moy = 8 [Etendue = 69-0] écart-type = 14 Q1 = 1 médiane=3 et Q3= 8
Scolarisation	Moy = 103 [Etendue = 144-66] écart-type = 13 Q1 = 98 médiane=102 et Q3= 111
Eau	Moy = 85 [Etendue = 100-36] écart-type = 16 Q1 = 75 médiane=93 et Q3= 97
Espérance	Moy = 69.3 [Etendue = 79.9-58.6] écart-type = 6.0 Q1 = 64.5 médiane=70.3 et Q3= 73.5
ⁱ Moy = Mean [Etendue = Maximum-Minimum] écart-type = SD Q1 = 25% médiane=Median et Q3= 75%	