

Master 1 : Économétrie et statistique
Analyse de données et descriptive

**Analyse du taux de mortalité chez les enfants de 0 à 5 ans
dans les pays en développement.**

Sommaire

Sommaire.....	2
I. Introduction.....	3
II. Analyse détaillée du sujet.....	5
III. Analyse en Composantes Principales ACP	21
IV. Corrélation entre la variable à expliquer et les variables latentes	40
V. Conclusion et discussion des résultats	47
VI. Annexes.....	48
VII. Bibliographie	63

I. Introduction

La mortalité infanto-juvénile, c'est-à-dire la mortalité des enfants de 0 à 5 ans, connaît une baisse importante depuis les années 1990. En effet, au niveau mondial, le taux de mortalité a diminué, le nombre de décès « *passant de 93 à 38 pour 1 000 naissances vivantes entre 1990 et 2019* »¹. Cependant, la mortalité infanto-juvénile reste encore élevée dans certaines régions, telles que l'Afrique subsaharienne qui détient le taux de mortalité infanto-juvénile le plus élevé au monde « *avec un enfant sur 13 qui meurt avant son cinquième anniversaire* »¹. De plus, « *la moitié des décès de moins de 5 ans en 2019 se sont produits dans seulement 5 pays : le Nigeria, l'Inde, le Pakistan, la République démocratique du Congo et l'Éthiopie* »¹. C'est dans le but de diminuer la mortalité dans les pays où elle est encore élevée que l'Organisation des Nations Unies au travers des objectifs de développement durable souhaite « *éliminer les décès évitables de nouveau-nés et d'enfants de moins de 5 ans d'ici à 2030* »². En effet, l'accès aux services de santé pour tous adultes ou enfants est un objectif présent dans tous les pays. L'une des problématiques étant d'éviter les difficultés financières à la suite d'une consultation médicale : « *chaque année un nombre considérable de familles sont plongées dans la pauvreté en raison des dépenses de santé qu'elles doivent régler directement* »³.

C'est au regard de ces chiffres clés, que nous avons voulu approfondir nos recherches sur le sujet. Nous nous sommes posées la question suivante : qu'est-ce qui explique cette disparité importante entre les différents pays en termes de mortalité infanto-juvénile ? Réaliser une analyse des facteurs pouvant expliquer ces taux de mortalité élevés permet de comprendre les politiques publiques prises dans le but de diminuer le taux de mortalité.

¹<https://www.who.int/fr/news-room/fact-sheets/detail/children-reducing-mortality>

²<https://www.unwomen.org/fr/news/in-focus/women-and-the-sdgs/sdg-3-good-health-well-being#:~:text=D'ici%20%C3%A0%202030%2C%20%C3%A9liminer,000%20naissances%20vivantes%20au%20plus.>

³<https://www.who.int/fr/news/item/13-12-2017-world-bank-and-who-half-the-world-lacks-access-to-essential-health-services-100-million-still-pushed-into-extreme-poverty-because-of-health-expenses>

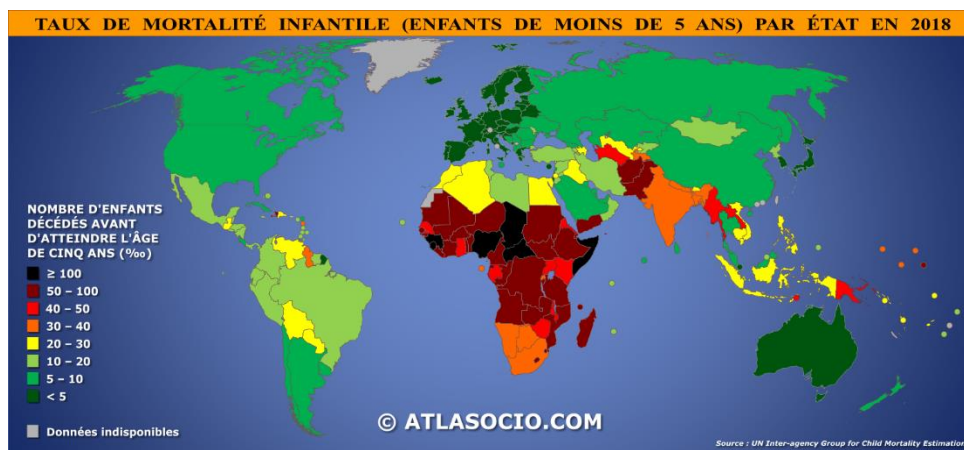
Pour y répondre, nous sommes parties de la littérature afin de trouver dans un premier temps des facteurs pouvant influencer la mortalité des enfants de 0 à 5 ans. Dans un second temps, nous développerons le contexte de notre analyse. Ensuite, nous utiliserons différents outils afin d'avoir une analyse globale tels que l'analyse en composantes principales ainsi qu'une régression linéaire multiple. Cela nous permettra de mettre en évidence les facteurs qui impactent ou non de manière positive ou négative la mortalité infanto-juvénile.

II. Analyse détaillée du sujet

A. Pertinence de la variable à expliquer

La mortalité infanto-juvénile correspond au nombre de décès survenus chez les enfants de moins de cinq ans par rapport à l'ensemble des enfants de cette tranche d'âge. C'est un sujet capital, faisant partie des objectifs au développement durable car « *plus de 60 % de ces décès sont considérés comme évitables par l'ONU grâce à des mesures peu coûteuses* »⁴. Nous pouvons constater sur la figure n°1 que la mortalité infantile est plus faible dans les pays développés avec moins de 5 décès en 2018 que dans les pays en développement comptabilisant plus de 100 décès avant l'âge de cinq ans. En 2018, la Somalie ou le Nigéria faisaient partie des pays avec le plus haut taux de mortalité des enfants de moins de 5 ans. Ainsi, le taux de mortalité des enfants de moins de 5 ans sera notre variable illustrative.

Figure n° 1 : Carte exposant le taux de mortalité infanto-juvénile dans le monde



<https://atlasocio.com/cartes/recherche/selection/taux-mortalite-infantile-moins-5-ans.php>

Nous venons de voir les chiffres clés de la mortalité infanto-juvénile représentée sur une carte. Cependant, afin d'avoir une étude sur les données les plus récente possible nous avons utilisé les données fournies par la banque mondiale et l'OMS. Notre analyse se fera alors sur la mortalité des enfants de moins de 5 ans en 2020. Afin de pouvoir faire une analyse complète, nous avons dû garder uniquement les pays dont la variable à expliquer et les variables explicatives étaient connues. C'est pour cela que nous nous retrouvons avec un total de 62 pays, tous en développement, ce qui nous permet de nous focaliser sur

⁴<https://atlasocio.com/classements/demographie/mortalite/classement-etats-par-taux-de-mortalite-infantile-moins-de-5-ans-monde.php>

cette catégorie de pays. Dans notre base de données, la mortalité infanto-juvénile est exprimée pour 1000 naissances vivantes.

Nous allons maintenant vous présenter les différentes variables explicatives de ce phénomène

B. Les variables explicatives et leur corrélation avec la variable illustrative.

1. Les variables explicatives

Pour trouver les variables explicatives, nous avons tout d'abord orienté nos recherches sur les causes de la mortalité infanto-juvénile. Les causes sont principalement inscrites sur le thème de la santé. Nous avons cherché à expliquer la mortalité infantile par des facteurs pouvant être récurrents dans les pays en développement. Le problème de nutrition revient fréquemment dans l'explication du taux de mortalité des pays en développement. Ainsi, nous allons inclure la malnutrition infantile dans notre liste de variables. En effet, l'Institut Pasteur et L'OMS stipulent respectivement dans leurs articles que « Celle-ci est en cause dans près de la moitié de la mortalité infantile »⁵ et « Les enfants souffrant de malnutrition, en particulier ceux souffrant de malnutrition aiguë sévère, courent un risque plus élevé d'être victimes des maladies de l'enfance courantes »⁶. Par conséquent, nous allons intégrer plusieurs maladies dans notre étude car ces maladies engagent un nombre considérable de décès. Cela est confirmé par l'OMS attestant que le risque de mortalité est réduit grâce aux traitements contre les maladies : « pouvant être évitées ou traitées au moyen d'interventions simples et peu coûteuses notamment la vaccination, une nutrition adéquate, de l'eau potable, ainsi qu'une alimentation et des soins de qualité fournis par un prestataire de santé »⁷. Nous avons donc choisi de prendre en

⁵<https://www.pasteur.fr/fr/institut-pasteur/institut-pasteur-monde/programmes-internationaux/projet-afribiota-malnutrition-infantile#:~:text=Lutter%20contre%20la%20malnutrition%20infantile,-Un%20enfant%20sur&text=Celle%2Dci%20est%20en%20cause,du%20d%C3%A9veloppement%20physique%20et%20mental>

⁶<https://www.who.int/fr/news-room/fact-sheets/detail/children-reducing-mortality#:~:text=Les%20enfants%20souffrant%20de%20malnutrition,la%20pneumonie%20et%20le%20paludisme.>

considération des variables de santé telles que le taux de vaccination du DTC3 (diphtérie-tétanos-coqueluche) ou la prévalence de l'hépatite B. Ces deux maladies peuvent être évitées par la vaccination des enfants.

Par la suite, nous avons élargi nos recherches sur l'environnement, en effet celui peut avoir un effet négatif sur la santé. Nous avons alors cherché des variables ayant probablement un impact sur la mortalité. La qualité de l'eau peut, par exemple entraîner des conséquences négatives sur la santé des enfants en provoquant des maladies comme la diarrhée. D'autres variables environnementales peuvent aussi jouer sur la mortalité. En effet, selon l'OMS« *On a estimé en 2012 que 26 % des décès infantiles et 25 % de la charge totale de morbidité des enfants de moins de cinq ans pourraient être évités par la réduction des risques environnementaux tels que la pollution de l'air, l'insalubrité de l'eau, les mauvaises conditions d'hygiène et d'assainissement ou les produits chimiques* »⁷ .

Enfin, nous avons fini par chercher des variables liées au progrès social comme l'alphabétisation des femmes. En effet, celle-ci peut avoir un impact significatif sur la santé des enfants. L'Institut Pasteur vérifie ce constat en stipulant que « *la mortalité infantile pourrait être réduite de 15 % si toutes les mères avaient une instruction primaire, et de 49 % si elles avaient reçu une éducation secondaire.* »⁸Cela peut s'expliquer car l'alphabétisation permet globalement d'acquérir une meilleure compréhension des informations données à l'orale ou à l'écrit. C'est en adéquation avec le fait que « *l'instruction permet aux femmes de connaître les bonnes pratiques d'hygiène, de protection contre certaines maladies (sida, paludisme, malaria...), de bien administrer les médicaments, de faire appel à des sage-femme pour le suivi des grossesses* »⁹.

En résumé, nos investigations nous ont montré que les facteurs de la mortalité proviennent principalement de l'hygiène de vie des pays et de l'ignorance de ce phénomène. Par conséquent, nous avons aussi pris en compte l'espérance de vie à la naissance car c'est « *l'un des indicateurs de l'état de santé les plus fréquemment utilisés. L'augmentation de l'espérance de vie à la naissance peut être attribuée à plusieurs facteurs,*

⁷<https://www.who.int/fr/publications-detail/WHO-FWC-IHE-17.01>

⁸<https://www.la-croix.com/Actualite/Monde/L-alphabetisation-des-femmes-un-facteur-de-developpement-2014-09-08-1202867>

dont l'élévation du niveau de vie, l'amélioration du mode de vie et l'augmentation du niveau d'instruction, ainsi qu'un accès plus large à des services de santé de qualité.»⁹

Ainsi, nos recherches littéraires nous ont montré qu'il existait de multiples facteurs expliquant la mortalité infanto-juvénile. Les variables que nous allons utiliser dans notre modèle vont être décrites par la suite. Nous avons principalement fait appel à la base de données intitulée « Statistiques mondiales de la santé 2022 » disponible sur le site de l'Organisation Mondiale de la Santé¹⁰ (OMS) pour réaliser l'analyse en composantes principales. Elle sera détaillée dans le chapitre 3. Cependant, avant de débiter l'analyse, nous allons observer le comportement de la variable illustrative avec chaque variable active dans la sous-partie suivante.

2. Rôle de variables actives et corrélation avec la variable illustrative

Pour observer les différentes corrélations, nous avons choisi de faire un tableau récapitulant les coefficients de corrélation entre la variable de mortalité et les différentes variables actives. Par la suite, nous nous appuierons sur les graphiques présentant 2 variables à la fois afin d'expliquer d'une part la variable active et réaliser d'autre part, l'interprétation des coefficients de corrélation.

Nous disposons de 12 variables actives présentées ci-dessous :

Figure n° 2: Définition des variables

Variables actives	
X1	PIB/habitant
X2	Taux de la maigreur de l'enfant (en%)
X3	Prévalence de l'hépatite B (appelé le VHB) chez l'enfant
X4	Couverture vaccinale diphtérie-tétanos-coqueluche (DTC3) chez les enfants ayant 1 an (en%)

⁹<https://data.oecd.org/fr/healthstat/esperance-de-vie-a-la-naissance.htm>

¹⁰<https://data.who.int/fr/indicators/i/2322814>

X5	Prévalence de l'anémie chez les femmes en âge de procréer (15 à 49 ans) en %
X6	Taux de scolarisation des filles dans le primaire
X7	Densité de médecins pour 10 000 habitants (en%)
X8	Densité d'infirmiers et de personnels obstétriques pour 10 000 habitants (en%)
X9	Dépenses publiques pour la santé (en%)
X10	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
X11	Accès aux services d'eau de base (en% de la population)
X12	Espérance de vie à la naissance
Y (variable illustrative)	Taux de mortalité des enfants de – de 5 ans (en%)

Voici ci-dessous un tableau récapitulatif de la corrélation entre la mortalité infanto-juvénile et les 12 variables actives choisies :

Figure n° 3 : Coefficient de corrélation entre la mortalité et les variables actives

	V1
Prévalence Hépatite B	0.80
Anémie	0.72
Maigreur extrême	0.34
Recherche et santé	0.04
Scolarisation	-0.06
Vaccin DTC3	-0.44
PIB/hab	-0.48
Dépenses publiques	-0.51
Infirmiers et personnels obstétriques	-0.62
Médecins	-0.64
Eau	-0.80
Espérance de vie à la naissance	-0.91

Nous observons que la mortalité infanto-juvénile a une corrélation positive forte avec la prévalence de l'hépatite B et l'anémie (comprise entre 0,5 et 1) et faible avec la maigreur extrême (moins de 0,5).

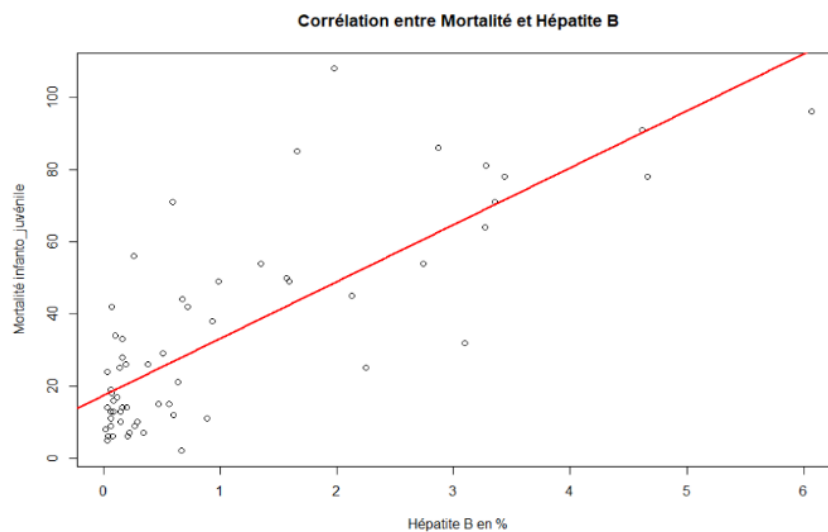
Inversement, la variable illustrative a une corrélation négative forte avec l'espérance de vie, l'eau, la densité de médecins, les infirmiers et les dépenses publiques (comprise entre -1 et -0.5), moyenne avec le PIB, la vaccination au DTC3 (de -0,5 à -0,3) puis faible avec le taux de scolarisation des filles dans le primaire (moins de -0,3).

Nous allons dans la suite décrire chaque variable active et sa relation avec la mortalité infantile.

a) Mortalité et Hépatite B

Nous avons choisi de prendre en considération la prévalence de l'antigène de surface l'hépatite B chez les enfants de moins de 5 ans. Celui-ci est le seul virus contre lequel il existe un vaccin. Cette infection entraîne une inflammation du foie dont environ 500 000 personnes décèdent par an.

Figure n° 4 : Graphique de corrélation entre Mortalité et Hépatite B



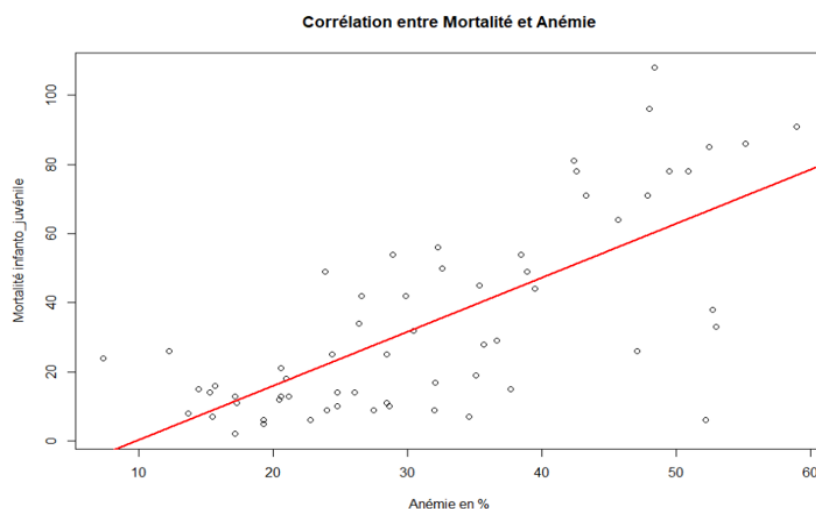
Ce graphique illustre une relation linéaire croissante entre les deux variables avec un fort coefficient de corrélation, en accord avec la figure n°3 qui nous indique un coefficient de 0,80. Autrement dit, la mortalité infanto-juvénile augmente considérablement lorsque la prévalence de l'Hépatite B, maladie transmissible le plus souvent à la naissance, augmente. Ce constat est en accord avec le fait que 2,5 millions

d'enfants de moins de 5 ans sont sauvés grâce à la vaccination contre les maladies infectieuses¹¹.

b) Mortalité et Anémie

Afin d'avoir un panel d'indicateurs de santé, nous avons décidé de prendre en compte la prévalence de l'anémie chez les femmes en âge de procréer (15 à 49 ans) en pourcentage. L'anémie étant un taux anormalement bas d'hémoglobine qui peut provenir de causes très diverses, allant des carences nutritionnelles aux affections gynécologiques et obstétricales. De plus, si l'anémie se présentait chez la femme enceinte, elle entraînerait un risque pour que « *le fœtus de ne reçoive pas la quantité d'oxygène suffisante pour sa croissance et son développement* »¹².

Figure n° 5 : Graphique de corrélation entre Mortalité et Anémie



Ce graphique illustre une relation linéaire croissante avec un fort coefficient de corrélation entre la mortalité des enfants de moins de cinq ans et la prévalence de l'anémie chez la femme de 15 à 49 ans. Autrement dit, la mortalité infanto-juvénile augmente de 0,72 (figure n° 3) lorsque la prévalence de l'anémie, augmente de 1. Ce constat est

¹¹<https://www.unicef.fr/convention-droits-enfants/sante/maladies-infantiles/hepatite-b/>

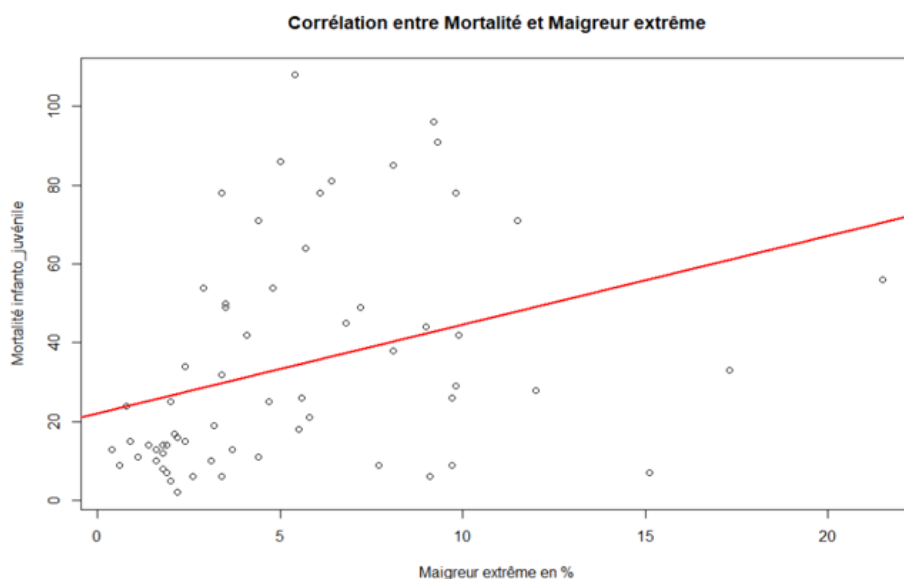
¹²<https://www.may.app/blog/grossesse/lanemie-pendant-la-grossesse-tout-savoir/#:~:text=Une%20an%C3%A9mie%20s%C3%A9v%C3%A8re%20n'est,essoufflement%20et%20de%20fatigue%20constante>

confirmé par l'article de l'OMS stipulant que « *Les enfants et les femmes enceintes sont particulièrement vulnérables, avec un risque accru de mortalité maternelle et infantile* »¹³.

c) Mortalité et maigreur extrême

Notre variable maigreur extrême correspond dans notre modèle à l'émaciation des enfants de moins de cinq ans. La malnutrition « *est en cause dans près de la moitié de la mortalité infantile* »¹⁴.

Figure n° 6 : Graphique de corrélation entre Mortalité et Maigreur extrême



Ce graphique nous montre une relation linéaire croissante entre les deux variables avec un coefficient de corrélation plus faible que précédemment (0,34). Cependant, la mortalité infanto-juvénile augmente lorsque le taux de la maigreur croît. Ceci est en accord avec la littérature.

d) Mortalité et Recherche et Santé

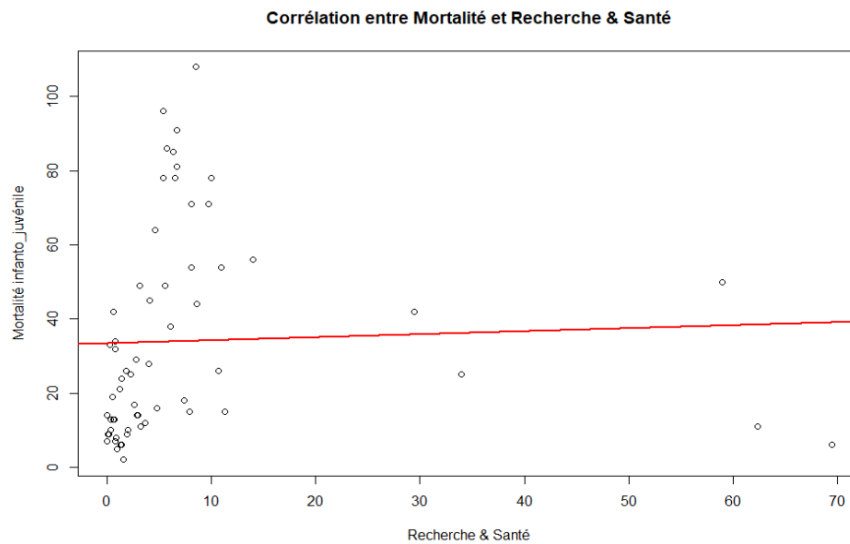
La variable Recherche et Santé représente l'aide publique au développement (APD) consacrée à la recherche en santé publique et aux secteurs de la santé de base, par habitant

¹³https://www.who.int/fr/health-topics/anaemia#tab=tab_2

¹⁴<https://www.pasteur.fr/fr/institut-pasteur/institut-pasteur-monde/programmes-internationaux/projet-afribiota-malnutrition-infantile>

aux pays bénéficiaire. D'après l'OCDE¹⁵, cette aide a pour objectif d'améliorer les conditions de vie ainsi que le développement économique des pays qui sont dans le besoin. Elle correspond au total des dépenses des donateurs. Nous avons utilisé cette variable car nous centralisons notre étude sur les pays en développement.

Figure n° 7 : Graphique de corrélation entre Mortalité et Recherche&Santé



Ce graphique nous expose une relation quasi inexistante avec un coefficient de corrélation proche de 0. Autrement dit, l'aide publique n'influence pas l'évolution du taux de mortalité.

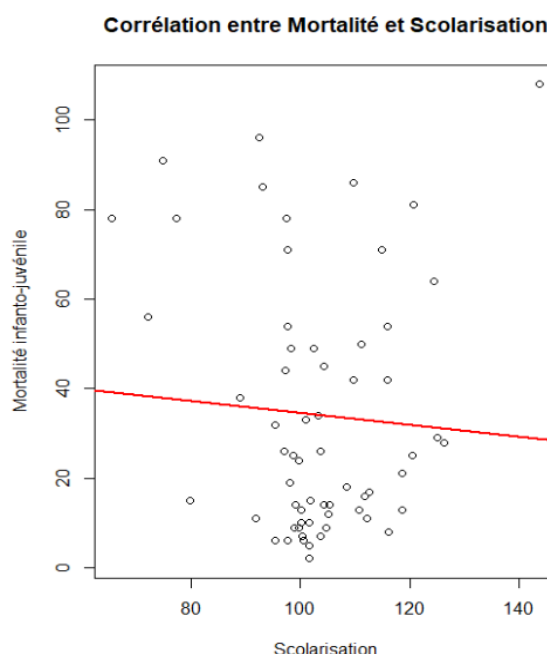
e) Mortalité et scolarisation

Notre variable scolarisation fait référence à la scolarisation des filles dans le primaire en pourcentage brut. Le taux brut se traduit par le « *rapport entre la scolarisation totale quel que soit l'âge et la population du groupe d'âge qui correspond officiellement au niveau d'éducation indiqué* »¹⁶. Cette variable est importante car la scolarisation bien qu'elle soit restreinte au niveau primaire dans notre modèle, permet aux enfants d'acquérir des compétences de base en lecture, écriture et mathématiques. Ainsi, cette instruction contribue à une meilleure compréhension globale.

¹⁵<https://www.oecd.org/fr/cad/financementpourledeveloppementdurable/normes-financement-developpement/aide-publique-au-developpement.htm>

¹⁶<https://donnees.banquemondiale.org/indicateur/SE.PRM.ENRR.FE>

Figure n° 8 : Graphique de corrélation entre Mortalité et Scolarisation



D'après ce graphique, la scolarisation des filles dans le primaire a une corrélation négative assez faible de -0,06 avec la mortalité infanto-juvénile. En d'autres termes, plus l'instruction des filles évolue de manière croissante, plus les risques diminuent. On peut l'expliquer par le fait que les femmes soient plus conscientes des risques liés à la grossesse et deviennent plus responsables. Ceci est confirmé par le journal de La Croix¹⁷.

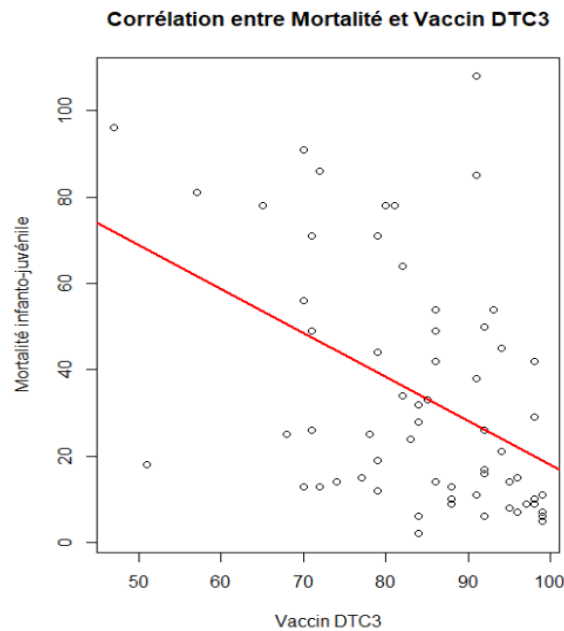
f) Mortalité et Vaccin DTC3

Le vaccin diphtérie-tétanos-coqueluche (DTC3) représente la couverture vaccinale des enfants de moins d'un an dans notre analyse. Ce dernier protège les enfants contre les maladies infectieuses les plus graves selon L'OMS¹⁸. La diphtérie étant une infection respiratoire pouvant entraîner la mort par asphyxie dans les cas les plus graves. En 2018, le tétanos, une maladie infectieuse aiguë a entraîné la mort de 25 000 nouveau-nés. La coqueluche est également une maladie respiratoire pouvant prendre une forme grave chez les femmes enceintes et les nourrissons. Le vaccin diminue fortement le nombre de cas.

¹⁷<https://www.la-croix.com/Actualite/Monde/L-alphabetisation-des-femmes-un-facteur-de-developpement-2014-09-08-1202867>

¹⁸<https://www.who.int/fr/news-room/fact-sheets/detail/immunization-coverage>

**Figure n° 9 : Graphique de corrélation entre
Mortalité et Vaccin DTC3**



La représentation graphique nous montre une corrélation négative entre les deux variables. En effet, plus la couverture vaccinale augmente, plus la mortalité décroît, le coefficient de corrélation étant de -0,44.

g) Mortalité et PIB par habitant

Le produit intérieur brut (PIB)¹⁹ est le principal agrégat mesurant l'activité économique. L'INSEE le définit comme étant « *la somme des valeurs ajoutées brutes nouvellement créées par les unités productrices résidentes une année donnée, évaluées au prix du marché* »²⁰ et Statista²¹ informe que le PIB par habitant est mesuré par le rapport entre la richesse du pays et sa population.

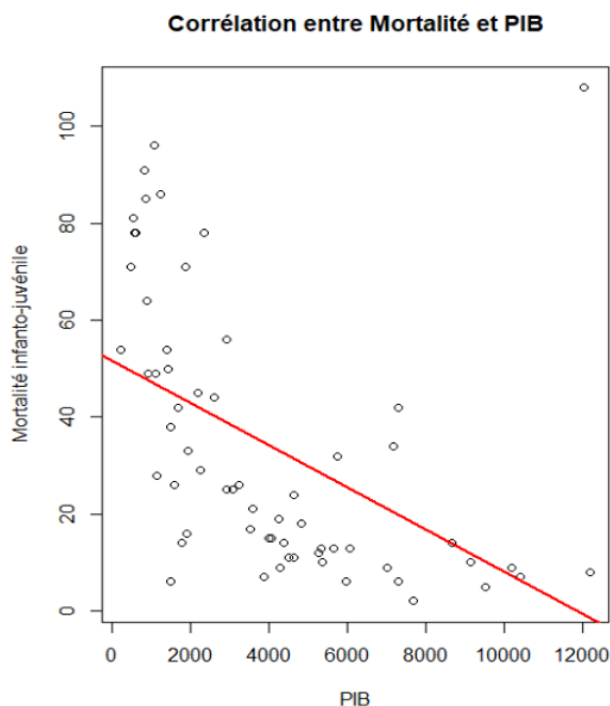
¹⁹<https://donnees.banquemondiale.org/indicateur/NY.GDP.PCAP.CD>

²⁰<https://www.insee.fr/fr/statistiques/2415846>

²¹<https://fr.statista.com/statistiques/661077/classement-pays-plus-eleve-monde-pib/#:~:text=Le%20PIB%20par%20habitant%20est,%C3%A9lev%C3%A9%20au%20monde%20en%202023%20>

La figure n°10 nous montre une corrélation négative entre les deux variables. En effet, lorsque le PIB par habitant augmente de 1 \$, la mortalité des enfants de moins de ans diminue de 0,48. Ce constat est confirmé par la littérature.

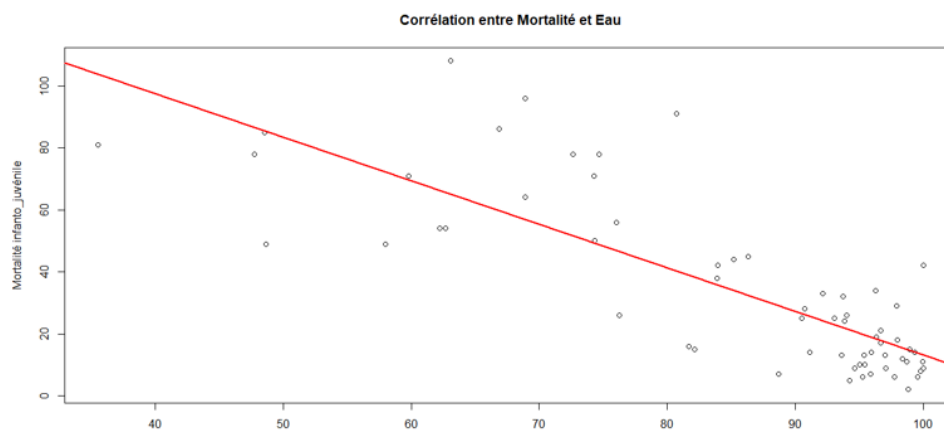
Figure n° 10 : Graphique de corrélation entre Mortalité et PIB



h) Mortalité et Eau :

La variable Eau²² correspond à l'accès aux services d'approvisionnement en Eau, d'assainissement et d'hygiène (EAH) adaptés dans les pays. L'accès à l'eau est un facteur

Figure n° 11 : Graphique de corrélation entre Mortalité et Eau



²²<https://donnees.banquemondiale.org/indicateur/SH.H2O.BASW.ZS>

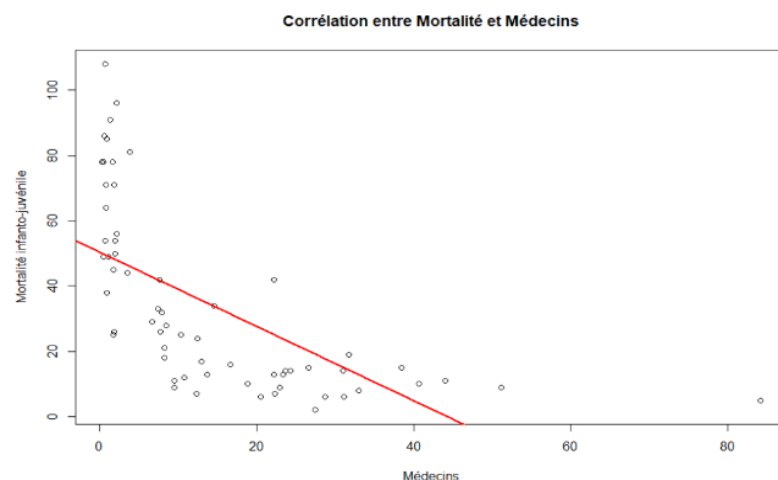
important afin d'éviter des problèmes comme la déshydratation ou la consommation d'eau contaminée qui peut transmettre de nombreuses maladies telles que le choléra ou la diarrhée. (Figure 11 ci-dessous)

Sur le graphique, nous pouvons observer une importante corrélation négative entre les services EAH et la mortalité avec un coefficient de corrélation de -0,8. En d'autres termes, plus la part des services permettant un accès à l'eau potable est élevée, plus la mortalité infanto-juvénile diminue. Cette observation est appuyée par l'article de (l'UNICEF) affirmant que « *Plus de 700 enfants de moins de 5 ans meurent chaque jour de maladies diarrhéiques faute d'accès à des services d'EAH adaptés* »²³ et du rapport de l'OMS.

i) Mortalité et densité de médecins :

La densité de médecins correspond à la quantité de médecins présents pour une population de 10 000 habitants. Nous avons choisi de prendre en compte ces données car elles sont directement liées à l'accès aux services de santé de la population. De plus, avec un accès plus important, les décès maternels diminuent.

Figure n° 12 : Graphique de corrélation entre Mortalité et Médecins



Le graphique illustre une corrélation négative entre les deux variables. En effet, lorsque la densité de médecins croît de 0,64 la mortalité connaît une diminution de 1 unité. Ces résultats corroborent l'article de l'OMS.²⁴

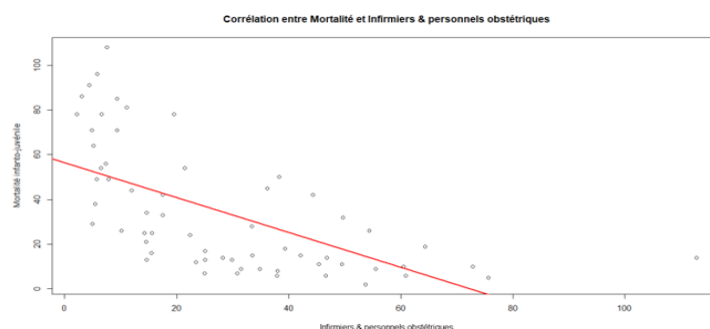
²³<https://www.unicef.org/fr/eau-assainissement-hygiene>

²⁴<https://www.who.int/fr/news-room/fact-sheets/detail/children-reducing-mortality>

j) Mortalité et infirmiers et personnels obstétriques :

La densité d'infirmiers et de personnels obstétriques traduit la quantité de personnels présents pour une population de 10 000 habitants. Ce paramètre est un indicateur complétant le précédent. En effet, « seuls 68% des accouchements se font accompagné de personnel médical dans les pays à faible revenu »²⁵ contre 99% dans les pays à revenu élevé.

Figure n°13 : Graphique de corrélation entre Mortalité et infirmiers



Le graphique illustre une corrélation négative entre les deux variables. En effet, plus la densité de personnels médicaux croît, plus la mortalité connaît une diminution. Ces résultats corroborent l'article de l'OMS.²⁶

k) Mortalité et Dépenses publiques :

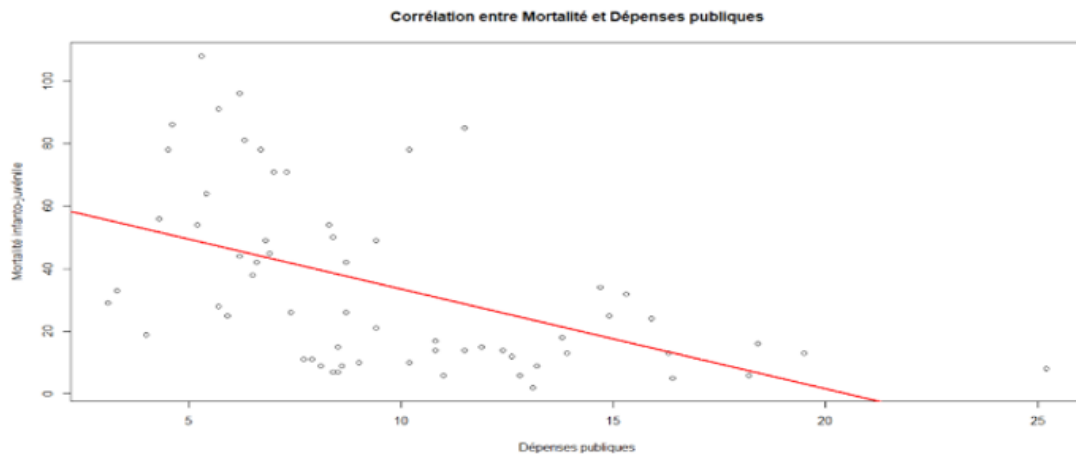
Les dépenses publiques correspondent aux dépenses publiques intérieures de santé en pourcentage des dépenses publiques générales de la santé c'est-à-dire la « *part des dépenses des administrations publiques finançant les dépenses courantes de santé* ». ²⁷ Les dépenses en santé prennent en compte la consommation de soins et de biens médicaux. (Figure 14 à la page suivante)

²⁵<https://www.who.int/fr/news-room/fact-sheets/detail/maternal-mortality>

²⁶<https://www.who.int/fr/news-room/fact-sheets/detail/children-reducing-mortality#:~:text=%C3%80%20l'%C3%A9chelle%20mondiale%2C%20les,de%20moins%20de%205%20ans>

²⁷<https://data.who.int/fr/indicators/i/B9C6C79>

Figure n° 14 : Graphique de corrélation entre Mortalité et Dépenses

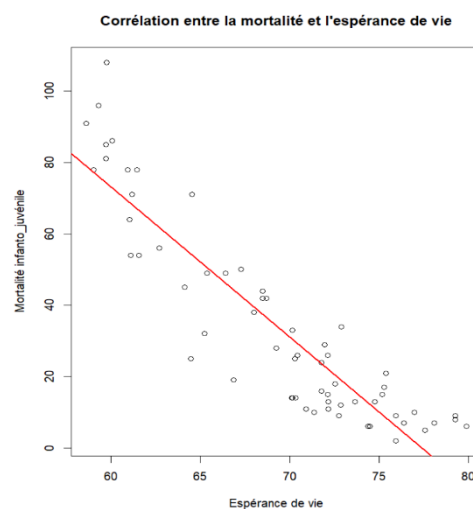


Le graphique illustre une corrélation négative entre les deux variables de coefficient de -0,51. En effet, plus la densité de personnels médicaux augmente, plus la mortalité connaît une diminution.

l) Mortalité et espérance de vie

L'espérance de vie²⁸ correspond à l'espérance de vie à la naissance c'est-à-dire le nombre d'années qu'il reste à vivre à un nouveau-né dans les conditions actuels de vie dans son pays. L'espérance de vie est un indicateur de la santé de la population ainsi que de la qualité de vie dans un pays.

Figure n° 15 : Graphique de corrélation



²⁸https://donnees.banquemondiale.org/indicateur/SP.DYN.LE00.IN?most_recent_year_desc=true

Sur la figure n°15 on peut voir que la corrélation entre la mortalité des enfants des moins de 5 ans et l'espérance de vie est négative. En effet, le coefficient de corrélation de -0,91.

III. Analyse en Composantes Principales ACP

A. Analyse statistique des variables

Dans ce chapitre, nous allons nous concentrer sur la quatrième étape de l'élaboration du modèle, l'analyse de données. Celle-ci est apparue au début du vingtième siècle, définie comme étant un vaste domaine utilisant diverses méthodes statistiques descriptives dans le cadre multidimensionnel. Pendant les années 30, l'économiste et statisticien, Harold Hotteling a réalisé la première analyse en composantes principales, devenue l'outil le plus utilisé pour l'étude de variables quantitatives. Il l'a fondée en s'inspirant des idées de Spearman datant de 1904. Ce dernier a en effet été le premier à se concentrer sur l'étude des variables latentes en étudiant les tests psychologiques avec pour objectif de comprendre l'hétérogénéité des résultats des individus. En résumé, l'analyse en composantes principales est une analyse descriptive qui consiste à faire une étude factorielle. En d'autres termes, cette dernière a pour but de synthétiser de nombreuses variables en deux variables latentes ou synthétiques tout en conservant le plus possible une image fidèle de la réalité.

L'ACP nous permettra de mettre en lumière les variables ayant le plus d'impact sur la mortalité des enfants de moins de cinq ans dans les pays en développement. Cependant avant de faire l'ACP, nous ferons dans un premier temps une analyse de notre base de données afin d'en mesurer sa qualité. Nous commencerons par observer la présence ou non de valeurs atypiques en interprétant les boxplot. L'observation sera complétée de différents tests afin de nous guider dans notre choix de les garder ou non. Par la suite, nous effectuerons une analyse univariée de toutes les variables.

Figure n° 16 : Boxplot de la mortalité

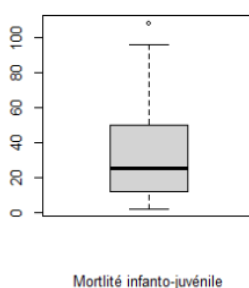
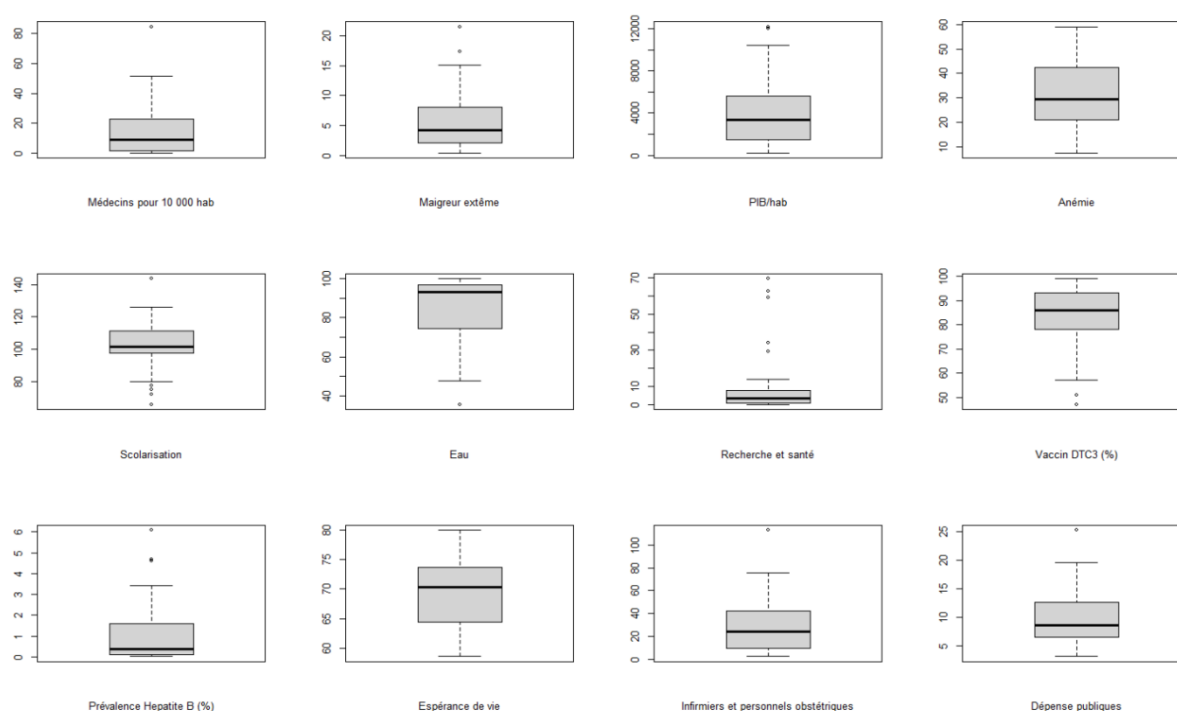


Figure n° 17 : Boxplot des variables actives



Nous avons fait des boxplot pour visualiser s'il y avait des valeurs atypiques pour chaque variable de notre base. Nous avons vérifié s'il y avait des valeurs atypiques pour éviter d'avoir une analyse biaisée par les valeurs extrêmes de l'échantillon. Comme on peut l'observer sur les boxplot des figures n°16 et n°17, plusieurs des variables semblent contenir des valeurs atypiques.

Nous avons dans un premier temps eu recours au test de Grubbs lorsqu'il y avait une seule valeur atypique et à celui de Rosner lorsqu'il y avait jusqu'à dix valeurs atypiques. Nous avons finalement trouvé 13 valeurs atypiques. Afin de savoir si nous conservions ou non ces valeurs, nous avons comparé le skewness et le kurtosis (annexe n°1) de la base de données avec et sans les pays possédant une valeur atypique. Nous avons choisi ces deux tests car le premier mesure l'asymétrie de la distribution et le deuxième examine la forme des queues de distribution. Avec les valeurs atypiques, le skewness vaut 0,94 et le kurtosis - 0,23 alors qu'en l'absence de valeurs atypiques, ils valent respectivement 1,15 et 0,58. Un skewness plus élevé signifie une asymétrie plus importante, et un kurtosis plus grand signifie une queue avec plus d'observations éloignées de la moyenne. L'augmentation des valeurs absolues que ce soit pour le skewness ou le kurtosis, nous a fait prendre la décision de garder les valeurs atypiques.

Cela signifie que la distribution est plus symétrique avec moins de valeurs éloignées de la moyenne. Ceci est plus pertinent pour une analyse de régression linéaire où la distribution doit être la plus normale possible. De plus, le fait de conserver toutes les valeurs nous permet de prendre en compte tous les pays dans notre analyse. Cela garantit une analyse complète sans exclusion des pays en développement, ce qui pourrait masquer des causes importantes de mortalité propres à ces pays.

Pour finir, l'analyse univariée permet d'appuyer nos hypothèses évoquées précédemment sur la forme de la distribution de nos données. Cette analyse consiste à observer la tendance centrale complétée par des mesures de dispersion.

A l'aide du logiciel R, nous obtenons le tableau suivant :

Figure n° 18 : Tableau de l'analyse statistiques des variables

Characteristic	N = 62 ¹
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
Maigreur 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
¹ Moy = Mean [Etendue = Maximum-Minimum] écart-type = SD Q1 = 25% médiane=Median et Q3= 75%	

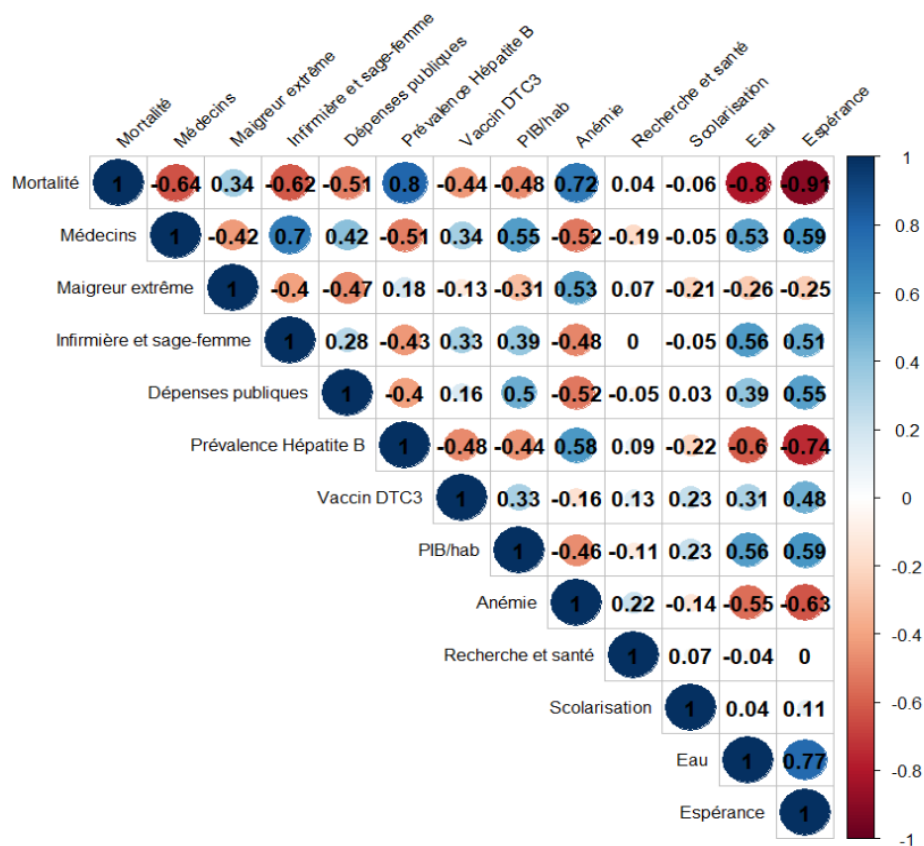
Nous pouvons remarquer que les paramètres de position, moyenne et médiane, sont assez proches l'une de l'autre pour la majorité des variables actives. Autrement dit, la distribution des valeurs est relativement symétrique dans sa globalité.

A propos de la variable illustrative « Mortalité », nous voyons que la médiane vaut 25. Cela signifie que 50% des pays ont un nombre de décès d'enfants de moins de 5 ans pour 1 000 naissances vivantes inférieur ou égal à 25 et 50% compte plus de 25 décès d'enfants âgés de moins de 5 ans, ce qui est éloigné de la moyenne. L'écart entre la

moyenne et la médiane s'explique par le fait qu'une minorité de pays font face à un taux de mortalité considérable, assez pour tirer la moyenne vers le haut. Ceci est confirmé par la présence des valeurs atypiques vues précédemment. Par conséquent, nous pouvons en déduire que la distribution des taux de mortalité est modérément disparate.

Nous pouvons compléter notre étude avec l'observation des quartiles de notre variable illustrative. La boîte centrale délimitée par le Q1 et le Q3 nous indique que 50% des pays comptent entre 12 et 50 décès d'enfants de moins de 5 ans pour 1 000 naissances vivantes. Le troisième quartile signifie que seulement 25% des pays dénombrent plus de 50 décès. Enfin, la moyenne valant 34 décès est donc inférieure au troisième quartile.

Figure n° 19 : Matrice des corrélations



Cette matrice des corrélations (figure n°19) nous donne un aperçu plus visuel sur les différentes relations qu'ont les variables entre elles. En effet, elle indique toutes les relations directes et indirectes existantes entre elles avec leur intensité.

Dans un premier temps, nous remarquons que la variable illustrative « Mortalité » a une corrélation négative forte ou modérée avec les variables de type médical telles que

« Médecins, infirmiers et personnels obstétricaux » ou encore le « Vaccin DTC3 ». Cette analyse est cohérente avec la littérature car les pays ayant accès aux soins sont plus aptes à proposer des traitements contre les maladies infectieuses. De plus, nous pouvons observer que le PIB est corrélé positivement avec ces dernières ainsi qu'avec la variable « Dépenses publiques ». En effet, les pays riches disposent de plus de ressources pour investir dans le secteur médical. Par ailleurs, nous observons aussi que l'espérance de vie et l'accès à l'eau sont fortement corrélés positivement. Ceci est compréhensible car l'eau est une ressource vitale pour notre organisme. Ainsi, les individus vivant dans un pays où la ressource en eau potable n'est pas menacée, peuvent escompter une espérance de vie plus élevée. Cette dernière est logiquement liée positivement aux variables relevant du médical. Par conséquent, la mortalité est corrélée positivement aux variables indiquant un manque de soins comme la maigreur, l'hépatite B ou encore l'anémie.

Ainsi, grâce à cette matrice, nous pouvons déjà conjecturer le sens des flèches que vont avoir les variables dans le cercle de corrélation. De plus, cette observation montre que notre base de données est en adéquation avec la littérature.

Maintenant que l'étude de notre base de données est accomplie, nous allons désormais procéder à l'analyse en composantes principales.

L'analyse en composantes principales suit une méthodologie composée de cinq étapes :

- Détermination des axes factoriels à l'aide des valeurs propres
- Analyse et interprétation des caractéristiques des variables
- Projection des variables par un cercle de corrélation
- Définition des variables latentes
- Projection des individus sur le plan factoriel

Dans l'ACP, nous excluons la variable illustrative. Cette dernière sera ajoutée lors de la projection sur le cercle pour appuyer notre interprétation réalisée au préalable. Par ailleurs, tous les calculs et visuels ont été réalisés par le logiciel R.

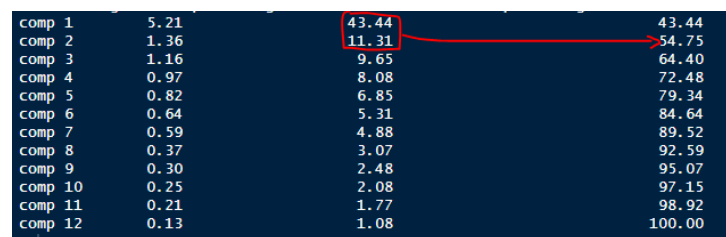
B. Valeurs propres et variables : contribution, corrélations, cosinus carrés

Dans cette section, nous allons commencer l'analyse par composante principale de notre base de données, cela nous permettra d'avoir une première visualisation de nos données avant de les projeter sur un cercle de corrélation.

1. Valeurs propres et nombre d'axes :

Pour commencer, nous allons identifier les axes qui prennent en charge l'inertie la plus importante. Cette dernière est calculée en faisant la somme des valeurs propres se trouvant sur la diagonale après la diagonalisation de la matrice des variances-covariances. La diagonalisation transforme les variances des variables en valeurs propres. Pour choisir les meilleurs axes du plan, nous prenons en compte le cumul le plus élevé du pourcentage d'inertie expliquée entre deux axes.

Figure n° 20 : Inertie de chaque axes



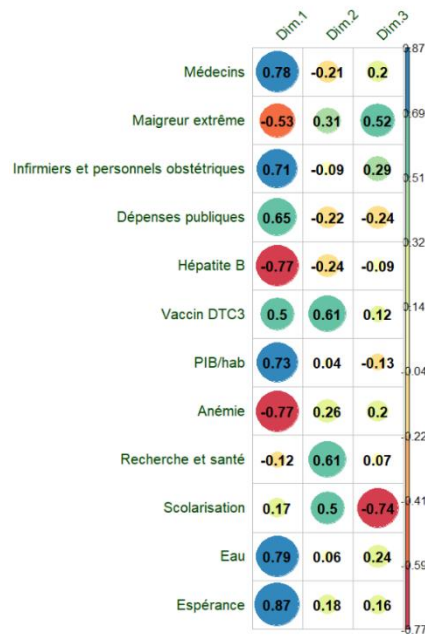
comp 1	5.21	43.44	43.44
comp 2	1.36	11.31	54.75
comp 3	1.16	9.65	64.40
comp 4	0.97	8.08	72.48
comp 5	0.82	6.85	79.34
comp 6	0.64	5.31	84.64
comp 7	0.59	4.88	89.52
comp 8	0.37	3.07	92.59
comp 9	0.30	2.48	95.07
comp 10	0.25	2.08	97.15
comp 11	0.21	1.77	98.92
comp 12	0.13	1.08	100.00

A l'aide de ce tableau, nous avons choisi de faire notre analyse sur le plan prenant en compte l'axe 1 et l'axe 2 car ils expliquent à eux seuls environ 55% de l'inertie, ce qui est très bon. Nous allons également réaliser l'analyse de notre modèle avec l'axe 3, explicitant une part d'inertie de 9,65%. De cette manière, notre modèle expliquera 64,40% de l'inertie grâce aux trois premiers axes.

2. Variables : coordonnées, contribution, cosinus carrés

Nous allons tout d'abord nous focaliser sur les coordonnées de chaque variable. Cette information va nous permettre de situer les variables dans les cadrans du cercle grâce au signe. Ici, les coordonnées sont assimilées aux coefficients de corrélation.

Figure n° 21 : Coordonnées des variables



D'une part, ce tableau nous montre que l'axe 1 est fortement corrélé positivement avec les variables relatives au monde médical, à l'eau, à l'espérance de vie et aux dépenses publiques. Par ailleurs, l'axe 1 est corrélé négativement avec les variables ayant une forte corrélation positive forte avec la variable illustrative comme vu précédemment dans la matrice des corrélations. D'autre part, l'axe 2 est fortement corrélé avec les variables définissant l'innovation ou le progrès social. Nous pouvons aussi remarquer que la variable « Vaccin DTC3 » est corrélée de façon importante avec les deux axes. Enfin, nous remarquons que l'axe 3 est fortement corrélé négativement avec la variable « Scolarisation ». Cette information nous incite à faire l'analyse sur les axes 1 et 3 en complément de celle impliquant l'axe 1 et l'axe 2.

Nous allons par la suite nous intéresser aux contributions et aux cosinus carrés, deux caractéristiques permettant de mesurer la qualité de la représentation du plan factoriel. Les contributions (figure n°22) permettent de mesurer la longueur des flèches et ainsi identifier les variables ayant une très forte influence sur les axes. Le cosinus carré (figure n°23) s'intéresse quant à lui à la qualité de représentativité des axes.

Figure n° 22 : Contributions des variables

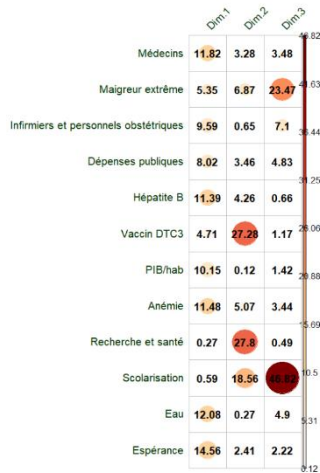
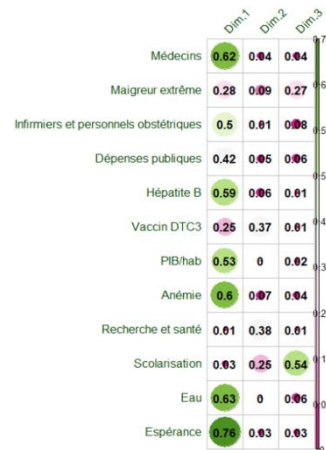


Figure n° 23 Cosinus carré des variables



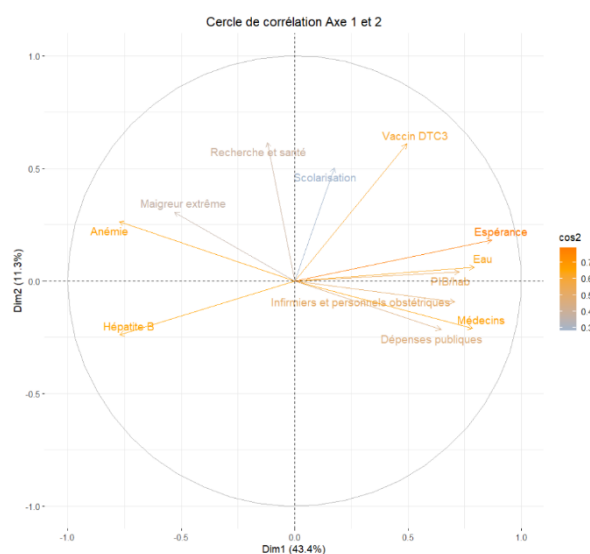
Nous constatons que les variables fortement corrélées avec l'axe 1 ont un cosinus carré proche de 1 et elles ont majoritairement une contribution oscillant aux environs de 12%. Par conséquent, les variables fortement corrélées à l'axe 1, ont une contribution forte ce qui signifie que l'axe 1 est très bien représenté par les variables de notre modèle. De plus, nous remarquons que la variable « Scolarisation » explique à elle seule près de 47% de l'inertie sur l'axe 3 et dispose du cosinus carré le plus élevé. Ainsi, l'axe 3 est principalement représenté par la variable « Scolarisation ». Cependant, nous rencontrons un problème pour l'axe 2. En effet, les trois variables, « Vaccin DTC3 », « Recherche et santé » ainsi que « Scolarisation » ayant les plus grandes contributions, ont toutes un cosinus carré allant de 0,38 à 0,25. Ce sont donc les variables représentant le plus l'axe 2 mais en comparaison avec l'axe 1 et 3 la corrélation entre l'axe 2 et les trois variables actives est moins forte.

C. Cercle de corrélation et variables latentes

1. Les cercles de corrélations

Nous allons maintenant passer à la projection des variables sur les cercles de corrélations. Cette étape permet de visualiser plus facilement les données précédentes, en effet, les coordonnées donnent le quadrant dans lequel se situe la variable, les contributions donnent la taille de la flèche et le cosinus carré l'angle avec les axes.

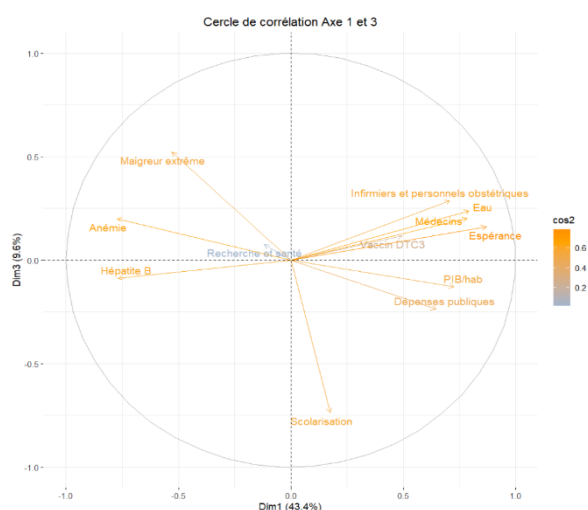
Figure n° 24 : Cercle de corrélation des axes 1 et 2



Le cercle confirme ce qui a été dit précédemment. Les variables fortement corrélées positivement avec l'axe 1 se trouvent à droite et celles corrélées négativement se situent à gauche. Par exemple, l'anémie est opposée à l'espérance de vie ou encore aux personnels hospitaliers. De plus, la variable « Vaccin DTC3 » se trouve en haut à droite car elle contribue le plus à l'axe 2 mais a un cosinus carré très faible.

Les variables bien représentées par l'axe 1 ont toutes une longue flèche proche de l'axe telle la variable « Espérance ». A la vue de ce cercle, nous pouvons envisager un premier regroupement de variables sur la base de leurs similitudes. Ce regroupement va nous permettre de définir les variables latentes.

Figure n° 25 : Cercle de corrélation des axes 1 et 3



Concernant l'axe 3, la variable « Scolarisation » caractérise à elle seule la composante. Sur le plan factoriel des axes 1 et 3, nous pouvons remarquer que les variables « Vaccin DTC3 » et « Recherche et santé » sont à l'épicentre et donc deviennent quasi inexistantes dans notre étude. Même si l'axe 1 n'a pas subi de grands changements en comparaison avec le plan factoriel précédent, nous pouvons remarquer que l'axe 3 a uniquement une forte corrélation négative avec la variable « Scolarisation ».

Pour conclure cette sous-partie, le travail réalisé précédemment nous a permis d'envisager des variables latentes grâce à l'analyse des caractéristiques des variables actives et de leur projection.

Dans la prochaine sous-partie, nous allons définir les trois axes, aussi appelés variables latentes, grâce aux variables qui représentent de façon plus ou moins importants chacun des axes. Cette étape est essentielle pour l'interprétation de notre analyse. En effet, après avoir défini les axes, nous projetterons les individus puis finirons par l'interprétation.

2. Définition des variables latentes

Nous allons à présent définir les variables latentes grâce à un tableau (à la page suivante) présentant la répartition des variables selon leur contribution par ordre décroissant et le signe de leur corrélation avec chaque composante.

Nous avons fait le choix de référencer les variables ayant une contribution supérieure ou égale à 5 et de préciser la qualité de leur représentativité.

Figure n° 26 : Tableau de la représentativité des axes

CORRÉLATION	POSITIVE	NÉGATIVE
Axe 1	Espérance Cos2=0,72 Eau Cos2=0,63 Médecins Cos2=0,62 PIB/hab Cos2=0,53 Infirmiers et personnels obstétriques Cos2=0,5 Dépenses publiques Cos2=0,42	Anémie Cos2=0,6 Hépatite B Cos2=0,59 Maigreux extrême Cos2=0,28
Axe 2	Recherche et santé Cos2=0,38 Vaccin DTC 3 Cos2=0,37 Scolarisation Cos2=0,25 Maigreux extrême Cos2=0,09	Hépatite B Cos2=0,06 Dépenses publiques Cos2=0,06 Médecins Cos2=0,04
Axe 3	Maigreux extrême Cos2=0,27 Infirmiers et personnels obstétriques Cos2=0,08	Scolarisation Cos2=0,54

Ainsi, nous pouvons constater que les variables faisant référence au corps médical (« Médecins » et « Infirmiers et personnels obstétriques »), à l'hygiène (« Eau ») à l'espérance de vie ainsi qu'à la richesse (« PIB/hab ») sont les plus contributives et très bien représentées avec un cosinus carré proche de 1 dans le sens positif sur l'axe 1. A l'opposé, nous retrouvons les variables liées aux maladies corrélées négativement avec cet axe. Cette variable latente représente donc la situation sanitaire des pays.

Concernant l'axe 2, nous pouvons dire qu'il est principalement représenté positivement par trois variables (« Recherche et Santé », « Vaccin DTC3 » et « Scolarisation »). Par conséquent, nous avons décidé de définir la variable latente de l'axe 2 comme étant le progrès social.

Enfin, l'axe 3 est décrit par plusieurs variables mais la scolarisation est la variable disposant de la plus forte contribution et du plus fort cosinus carré par rapport aux autres variables. Nous pouvons en effet le voir sur le cercle de corrélation des axes 1 et 3. Par ailleurs, les autres variables contribuent de façon assez similaire au deux axes ou de façon prononcée à l'axe 1 au détriment de l'axe 3. La variable latente correspond donc à la scolarisation.

De façon schématique nous pouvons résumer les variables latentes selon les 3 axes de la manière suivante :

Figure n° 27 : Schéma des variables latentes (axes 1 et 2)

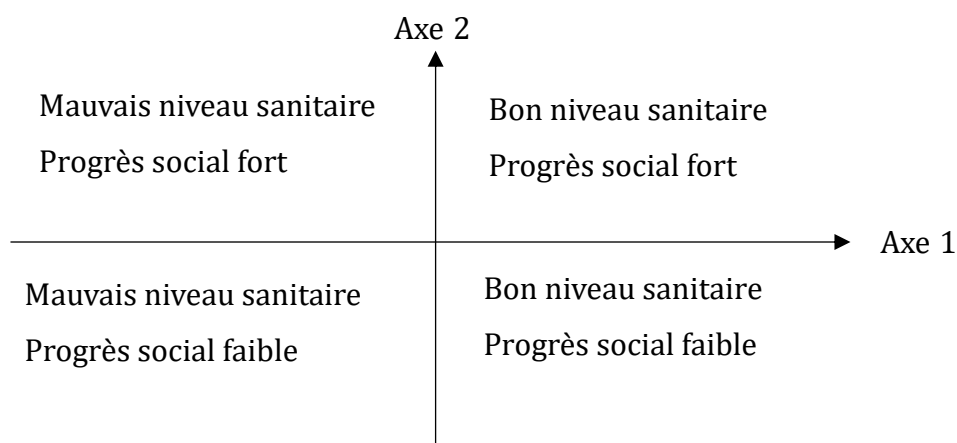
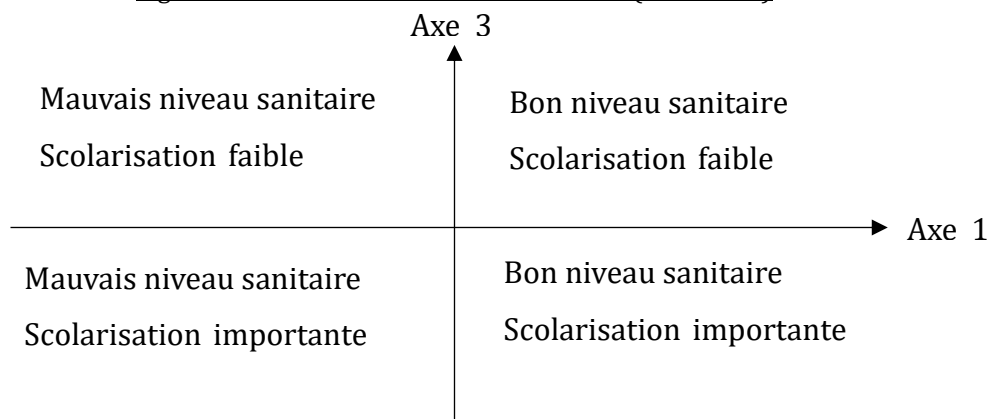


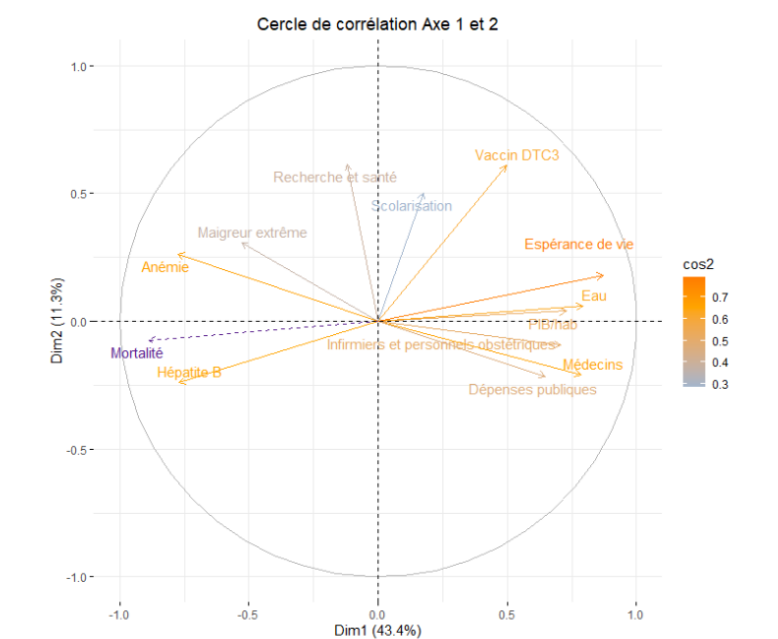
Figure n° 28: Schéma des variables latentes (axes 1 et 3)



3. Projection de la variable illustrative

Pour conclure sur l'étude des variables, nous allons projeter la variable illustrative sur les deux plans factoriels.

Figure n° 29 : Cercle de corrélation avec la mortalité



Sur le plan factoriel des axes 1 et 2, la variable illustrative « Mortalité » est bien à l’opposé des variables axées sur une bonne situation sanitaire. En d’autres termes, le niveau sanitaire et la mortalité sont corrélé négativement. Cela équivaut à dire que l’amélioration du niveau sanitaire conduit à la diminution de la mortalité des enfants de moins 5 ans comme stipulé dans l’article de l’OMS.²⁹

Par ailleurs, notre analyse est en accord avec la littérature concernant la malnutrition, représentée par la variable « Maigreur extrême ». En effet, les variables « mortalité » et « maigreur extrême » disposent de vecteurs allant dans le même sens. Le sens de ces vecteurs signifie que lorsque la malnutrition croît le long de l’axe 1, la mortalité augmente. Cela est appuyé par l’Unicef³⁰ et l’Institut pasteur³¹ affirmant respectivement que la malnutrition cause la mort d’un million et de trois millions d’enfants âgés de moins de 5 ans chaque année.

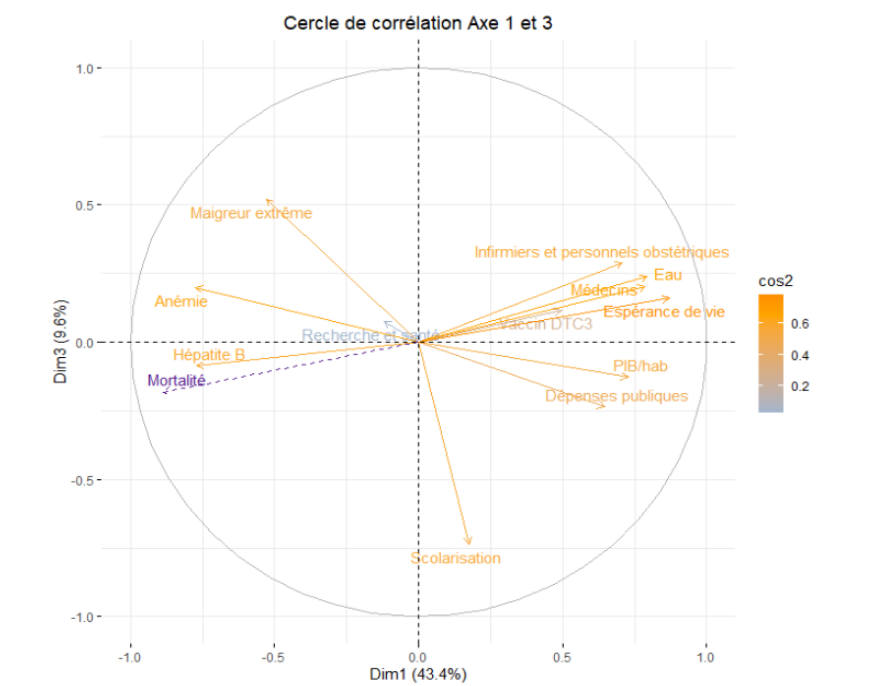
²⁹<https://www.who.int/fr/news-room/fact-sheets/detail/children-reducing-mortality#:~:text=%C3%80%20l'%C3%A9chelle%20mondiale%2C%20les,de%20moins%20de%205%20ans>

³⁰<https://www.unicef.fr/convention-droits-enfants/alimentation/malnutrition-infantile-et-des-enfants/>

³¹<https://www.pasteur.fr/fr/institut-pasteur/institut-pasteur-monde/programmes-internationaux/malnutrition-infantile>

Enfin, la « mortalité » se situe près de l'axe 1, la corrélation avec cette axe est donc plus forte qu'avec l'axe 2.

Figure n° 30 : Cercle de corrélation avec la mortalité



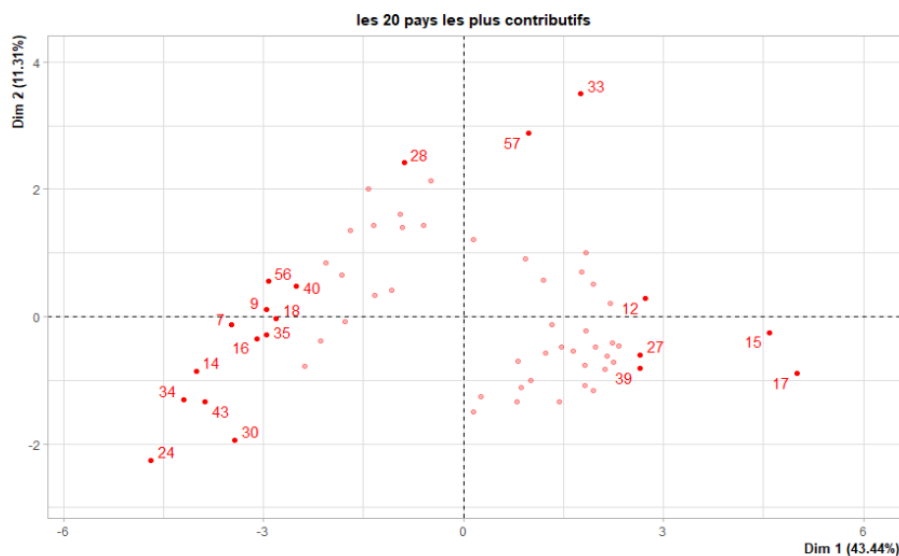
Ce plan factoriel, nous confirme l'analyse de l'axe 1 que l'on vient de faire. De plus, nous pouvons voir la corrélation négative entre la variable latente « Scolarisation » et la variable illustrative « Mortalité ». Cela signifie que, lorsque la scolarisation diminue le long de cet axe, la variable illustrative augmente. En effet, la « Mortalité » se situe dans la partie négative de l'axe 3, cela signifie que la corrélation entre les deux est négative. La « Scolarisation » diminuant lorsque l'on se déplace vers le haut, la mortalité des enfants de moins de cinq ans augmente.

D. Projection des individus sur le plan factoriel, interprétation

Maintenant que l'étude des variables est terminée, nous allons conclure ce chapitre par la projection des individus. Cette dernière va nous permettre de vérifier si la définition de nos variables latentes est en adéquation avec nos individus.

L'analyse des individus repose sur le même principe que celle des variables. En d'autres termes, nous allons projeter nos individus sur les deux plans factoriels. Cela nous permet de savoir où se situe chaque individu en fonction des variables latentes. Afin d'obtenir une interprétation précise de la projection des individus sur chaque plan factoriel, nous allons uniquement projeter les individus les plus contributifs. Pour renforcer notre observation, nous allons nous appuyer sur un tableau présentant la répartition des pays les plus contributifs selon les quadrants et ce, pour chaque plan. Par ce biais, nous pourrions mieux visualiser et expliquer l'opposition des pays.

Figure n° 31 : Projection des individus sur les axes 1 et 2



Nous allons commencer par l'interprétation de la projection des individus sur les axes 1 et 2 se trouvant ci-dessous.

Dans un premier temps, ce graphique nous montre très nettement l'opposition d'un gros groupe de pays avec un autre bloc de pays sur l'axe 1. Pour comprendre ce choix de répartition pris par le logiciel R, nous allons nous appuyer sur le tableau ci-dessous.

Figure n° 32 : Tableau récapitulatif des pays selon les quadrants de l'axe 1 et 2

Quadrant II :	Quadrant I :
40 : Mozambique 56 : Togo 28 : Kiribati	12 : Chine 57 : Tonga 33 : Maldives
Quadrant III :	Quadrant IV:
35 : Mauritanie 16 : Côte d'Ivoire 14 : Congo (démocratique)	15 : Costa Rica 39 : Monténégro 27 : Kazakhstan

Pour rappel, les composantes principales 1 et 2 correspondent respectivement à la situation sanitaire des pays et au progrès social. Prenons l'exemple du Monténégro qui se trouve dans le quadrant IV. Cela signifie que le Monténégro aurait une bonne situation sanitaire tout en ayant un progrès social faible.

Après consultation de nos données, nous pouvons en effet dire que ce pays dispose d'un corps médical bien présent avec un nombre relativement important de médecins, d'infirmiers et de personnels obstétricaux avec une espérance de vie élevée (75 ans). De plus, les chiffres concernant les maladies telles que l'anémie et l'hépatite B sont très faibles. En effet, les taux sont respectivement aux alentours de 17% et de moins d'1%. Enfin, le taux de mortalité des enfants âgés de moins de 5 ans valant 2 pour 1 000 naissances vivantes est très faible. Cependant, la recherche spécialisée dans la santé n'est pas présente au Monténégro. En effet, le pays bénéficie d'une très faible aide publique au développement destinée aux secteurs de la recherche médicale et de la santé de base oscillant les 1\$. Ainsi, les données du Monténégro sont en adéquation avec la définition de nos variables latentes.

De manière synthétique, les pays ayant des similitudes avec le Monténégro sont proches de celui-ci comme le Kazakhstan qui bénéficie d'un excellent corps médical mais bénéficiant d'une aide quasi inexistante dans la recherche. Toutefois, nous remarquons

que les pays Costa Rica (individu 15) et Cuba (individu 17) sont très éloignés. Cet éloignement s'explique par une espérance de vie très élevée atteignant presque 80 ans.

Dans le quadrant III, nous retrouvons les pays opposant le Costa Rica. En effet, ces pays disposent d'une très mauvaise situation sanitaire avec un corps médical quasi inexistant et par conséquent, un fort taux de mortalité comme la Côte d'Ivoire avec 78 décès pour 1 000 naissances vivantes. De plus, dans le quadrant II, nous retrouvons les pays ayant un fort taux de mortalité et d'une aide relativement plus élevée que le Monténégro mais reste toujours faible, étant aux alentours de 4\$. Nous pouvons aussi remarquer que l'individu 28 correspondant à Kiribati est très éloigné des autres points du quadrant. Bien que ce pays ait un taux de mortalité relativement élevé (50 décès pour 1 000 nouveau-nés), il bénéficie d'une aide considérable allant jusqu'aux alentours de 60\$.

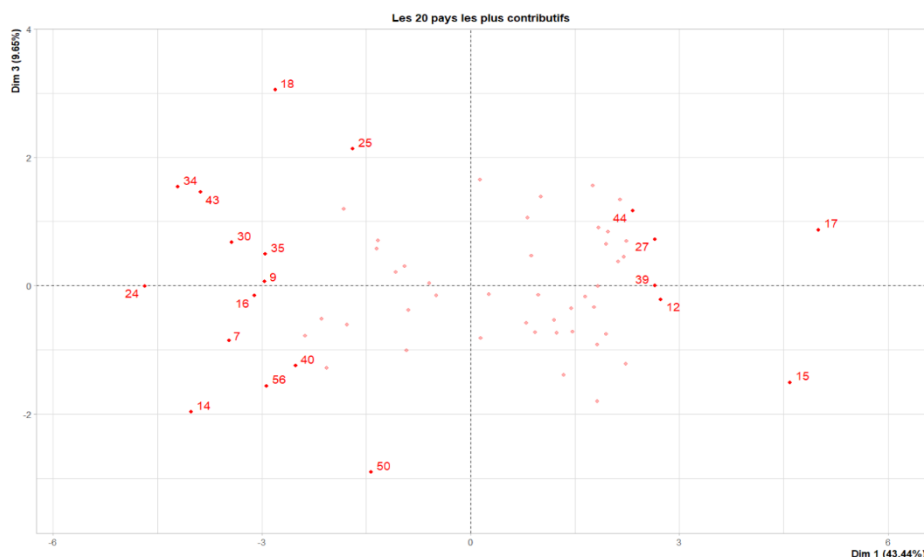
Enfin, nous pouvons remarquer que la Chine est très proche de l'axe 1 car elle bénéficie d'une aide quasi nulle. Cependant, le fait qu'elle dispose d'un très faible taux de mortalité et d'un corps médical relativement élevé explique son emplacement dans le quadrant I. Pour finir, les Maldives (individu 33) et le Tonga sont très écartés par rapport aux autres. En effet, l'écartement du Tonga s'explique par le fait que le pays bénéficie d'une aide conséquente de 62\$ dans la recherche, d'où sa proximité avec l'axe 2 et a le plus fort taux de mortalité comparé à ses camarades se trouvant dans le quadrant. En d'autres termes, ce pays a une situation relativement bonne avec un nombre élevé de personnels obstétriques mais un nombre trop faible de médecins.

Ainsi, nous pouvons confirmer d'une part que les pays confrontés à une bonne situation sanitaire se trouvent bien à droite de l'axe 1 et que d'autre part, les pays bénéficiant d'une faible aide dans la recherche consacrée à la santé se situent en bas de l'axe 2. Ce raisonnement est aussi applicable dans le sens inverse.

Nous allons par la suite procéder à l'interprétation du plan factoriel incluant les axes 1 et 3.

Tout d'abord, nous allons projeter comme précédemment les individus sur nos deux axes. Ensuite, nous interpréterons la position des pays les plus contributifs de notre modèle, affichés en gras sur le graphique ci-dessous.

Figure n° 33 : Projections des individus sur les axes 1 et 3



Afin d'identifier à quel pays chaque point fait référence, nous allons nous inspirer du tableau ci-après qui présente un extrait de pays contribuant le plus selon le quadrant dans lequel ils se situent.

Figure n° 34 : Tableau récapitulatif des pays selon les quadrants de l'axe 1 et 3

Quadrant II :	Quadrant I :
18 : Djibouti	17 : Cuba
26 : Jordanie	27 : Kazakhstan
34 : Mali	44 : Ouzbékistan
Quadrant III :	Quadrant IV:
14 : Congo (démocratique)	12 : Chine
50 : Sierra Leone	15 : Costa Rica
56 : Togo	

Nous pouvons maintenant analyser si le placement des pays dans le plan factoriel de l'axe 1 et 3 est en adéquation avec nos variables latentes. Pour commencer, nous pouvons prendre l'exemple de Cuba. Cuba se situe dans le quadrant II, cela signifie que ce pays devrait avoir une bonne situation sanitaire mais une scolarisation plutôt faible. En vérifiant les données de ce pays, nous pouvons constater qu'en effet, le nombre de médecins et d'infirmiers et personnels obstétricaux est important. De plus, l'anémie ou la maigreur extrême sont quant à elles, faibles. En ce qui concerne la scolarisation, Cuba fait partie des 50% les plus faibles de notre base de données. Sa position sur le plan factoriel est donc logique lorsque nous prenons en compte nos données.

Afin de s'assurer de la véracité de notre modèle, nous avons vérifié si les pays sont placés de manière logique avec notre plan. Tout comme pour Cuba, les pays semblent être placés au bon endroit. En effet, le Togo a un fort taux de scolarisation alors que le Mali en a un très faible. Ou encore, le Kazakhstan disposant d'un niveau sanitaire élevé explique sa position dans le quadrant I alors que Sierra Leone a un mauvais niveau sanitaire.

Nous pouvons donc conclure en disant que les pays avec un bon niveau sanitaire sont bien à droite de l'axe 1. Tandis que les pays avec une scolarisation faible se trouvent en haut de l'axe 3. Le corollaire est également valide.

Nous avons maintenant terminé notre analyse en composantes principales. Nous avons défini des variables latentes que nous avons précédemment vérifiées grâce à la projection des individus. Cela nous permet de passer à la régression linéaire multiple. Cette dernière nous permettra de confirmer nos interprétations faites dans l'ACP. En effet, nous connaissons le véritable lien de corrélation entre la mortalité infanto-juvénile et les trois variables latentes.

IV. Corrélation entre la variable à expliquer et les variables latentes

A. Régression linéaire multiples

Afin de réaliser la régression linéaire multiple, nous devons vérifier la conformité des hypothèses pour utiliser la méthode des moindres carrés ordinaires (MCO). Ces hypothèses doivent être acceptées pour les résidus du modèle estimé c'est-à-dire que le modèle doit suivre les conditions suivantes : les résidus doivent suivre une loi normale, l'homoscédasticité des erreurs doit être vérifiée ainsi que l'exogénéité des variables explicatives. Dans notre analyse, nous allons considérer que nos variables sont exogènes.

Figure n° 35 : Q-Q plot des résidus

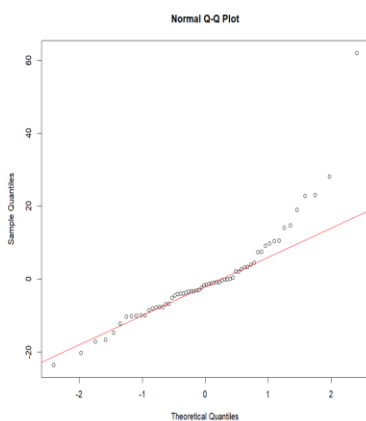


Figure n° 36 : Boxplot des résidus

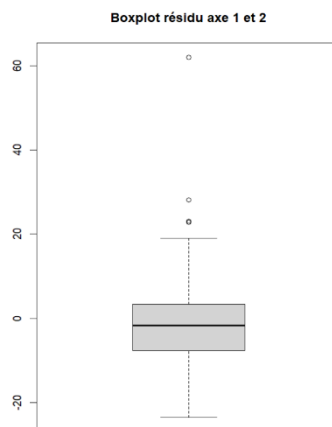
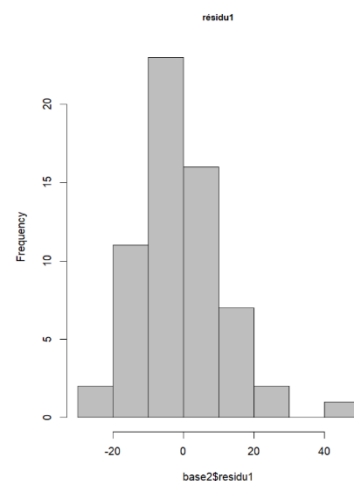


Figure n° 37 : histogrammes des résidus



Tout d'abord, nous souhaitons savoir si les résidus suivent une loi normale. Nous visualisons cette hypothèse avec le Q-Q plot. Au regard de ce graphique, nous pouvons penser que la distribution suit globalement une loi normale. Cependant, la normalité de la distribution n'est pas vérifiée avec le test paramétrique de Shapiro-Wilk (annexe n°4). Cela s'explique par le fait que le test ne soit pas optimal dû à la taille de notre échantillon qui est supérieure à 50 pays. De plus, cela peut s'expliquer par les valeurs élevées qui s'écartent de la ligne de normalité. En effet, il faut noter que sur la boxplot des résidus, nous avons trois valeurs aberrantes, ce qui peut influencer le résultat du test. Nous avons donc voulu vérifier la normalité de cette distribution avec un test non paramétrique qui est moins restrictif et plus adapté à la taille de notre échantillon. Le test Kolmogorov-Smirnov confirme alors la normalité de la distribution des résidus (figure n°38). En effet, la

p-value de ce test est supérieure à 0,05, la normalité de la distribution est donc acceptée au seuil de risque de 5%. Nous avons donc pris la décision de considérer la normalité de la distribution des résidus. De plus, on peut observer avec la boxplot des résidus que notre modèle a une distribution symétrique des résidus. Cependant nous avons trois valeurs aberrantes, nous avons décidé de les garder car l'impact de toutes les variables est important dans notre analyse. De plus, notre distribution suit une loi normale donc supprimer ces individus n'est pas obligatoire pour avoir une distribution qui suit la loi normale. L'histogramme confirme la symétrie de notre modèle ainsi que la présence de valeur critique à droite.

Figure n° 38 : Résultat du test Kolmogorov-Smirnov

```
> ks.test(base2$residu,"pnorm",mean(base2$residu),sd(base2$residu))

Asymptotic one-sample Kolmogorov-Smirnov test

data:  base2$residu
D = 0.16314, p-value = 0.07376
alternative hypothesis: two-sided
```

Pour pouvoir avoir un modèle interprétable par la régression linéaire, une autre des hypothèses qui doit être vérifiée est l'homoscédasticité des erreurs. L'hétéroscédasticité correspond à une variance des résidus qui n'est pas constante pour toutes les variables indépendantes. Nous pouvons vérifier cette hypothèse avec le test de Breusch-Pagan. Dans notre cas, l'homoscédasticité de notre modèle est validé (figure n°39). En effet, la p-value du test est supérieure à 0,05 ce qui signifie que H_0 n'est pas acceptée, l'hétéroscédasticité est donc refusée, notre modèle vérifie donc l'homoscédasticité des erreurs.

Figure n° 39 : Résultat du test Breusch-Pagan

```
> bptest(modele)

studentized Breusch-Pagan test

data:  modele
BP = 4.3537, df = 2, p-value = 0.1134
```

Nous avons des résultats similaires pour les résidus de l'axe 1 et 3 (annexe n° 5 à 9). Maintenant que nous avons vérifié les hypothèses de la distribution des résidus, nous pouvons faire une MCO pour notre régression linéaire multiple. Nous allons maintenant réaliser notre analyse économétrique.

Figure n° 40 : Résultat de la RLM de l'axe 1 et 2

```
Call:
lm(formula = base2$Mortalité ~ res.pca$ind$coord[, 1] + res.pca$ind$coord
    2])

Residuals:
    Min       1Q   Median       3Q      Max
-23.565  -7.461  -1.641   3.337  61.971

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    34.2258     1.6566  20.660  <2e-16 ***
res.pca$ind$coord[, 1] -10.7669     0.7256 -14.839  <2e-16 ***
res.pca$ind$coord[, 2]  -1.7657     1.4218  -1.242    0.219
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 13.04 on 59 degrees of freedom
Multiple R-squared:  0.7898,    Adjusted R-squared:  0.7827
F-statistic: 110.9 on 2 and 59 DF,  p-value: < 2.2e-16
```

Dans un premier temps, nous allons analyser notre modèle pour l'axe 1 et 2. Comme nous pouvons le voir sur la figure n°40 le modèle que nous avons choisi est significatif dans les axes 1 et 2 au seuil de risque de 5%. En effet, le test de Fisher (F-statistic) nous donne un résultat où la p-value est inférieur à 0,05, l'hypothèse nulle ou tous les coefficients sont égaux à zéro est refusée, cela signifie qu'au moins un des coefficients est significativement différent de zéro. Dans ce premier cas, le R carré ajusté est de 78%, cela signifie que 78% de la variabilité de notre modèle est expliquée. Nous pouvons aussi dire que l'axe 1 est significatif au seuil de risque de 5%. En effet, le test de Student donne une p-value inférieur à 0,05, le coefficient de l'axe 1 est donc significativement différent de zéro, l'axe 1 a alors un impact sur la mortalité infanto-juvénile. Son coefficient étant de -10,76, cela signifie que lorsque la situation sanitaire augmente d'une unité sur l'axe 1, la mortalité diminue de 10,76 unités. Cependant, le coefficient de l'axe 2 n'est pas significatif, nous ne pouvons donc pas l'interpréter.

Figure n° 41 : Résultat de la RLM de l'axe 1 et 3

```
Call:
lm(formula = base2$Mortalité ~ res.pca$ind$coord[, 1] + res.pca$ind$coord[,
  3])

Residuals:
    Min       1Q   Median       3Q      Max
-23.848  -8.200  -1.368   6.063  44.833

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)      34.2258     1.5449   22.154 < 2e-16 ***
res.pca$ind$coord[, 1] -10.7669     0.6767  -15.912 < 2e-16 ***
res.pca$ind$coord[, 3]  -4.6775     1.4358   -3.258  0.00186 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 12.16 on 59 degrees of freedom
Multiple R-squared:  0.8172,    Adjusted R-squared:  0.811
F-statistic: 131.9 on 2 and 59 DF,  p-value: < 2.2e-16
```

Tout comme pour l'axe 1 et 2, la figure n°41 nous montre que le modèle est significatif au seuil de risque de 5% pour les axes 1 et 3. En effet, le test de Fisher a une p-value inférieure à 0,05. Dans ce second cas, 81,1% de la variabilité du modèle est expliquée, le R^2 ajusté étant de 0,811. De plus, les axes 1 et 3 sont tous les deux significatifs au seuil de risque de 5%, la p-value du test de Student des deux axes étant inférieures à 0,05. Comme pour la régression linéaire multiple précédente, le coefficient de l'axe 1 est toujours de -10,76. Mais cette fois-ci, nous pouvons interpréter le coefficient de l'axe 3, qui vaut -4,67. Cela signifie que lorsque la scolarisation diminue d'une unité sur l'axe 3, la mortalité des enfants de moins de cinq ans augmente de 4,67.

En résumé, la régression linéaire de notre modèle montre un impact significatif des axes 1 et 3 sur la mortalité infanto-juvénile, alors que l'axe 2 n'a pas d'impact significatif sur la mortalité infanto-juvénile. De plus, pour chaque cas le modèle explique une grande partie de la variabilité de la mortalité, en effet les R-carrés ajustés sont élevés dans les deux cas.

B. Relation entre la variable illustrative observée et prédite par la RLM

Figure n° 42 : Modèle observé

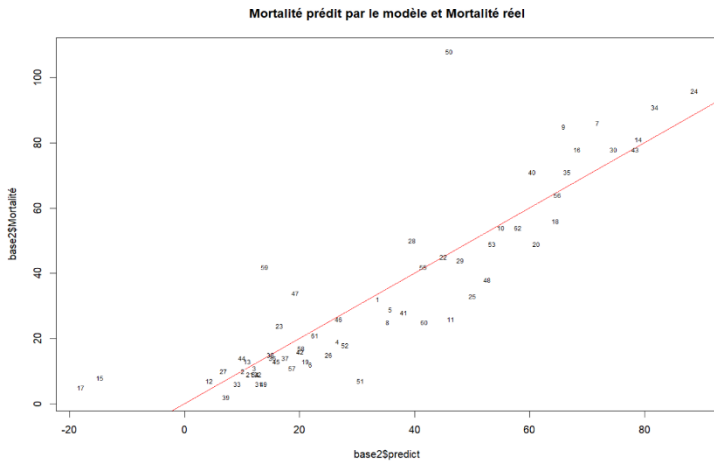
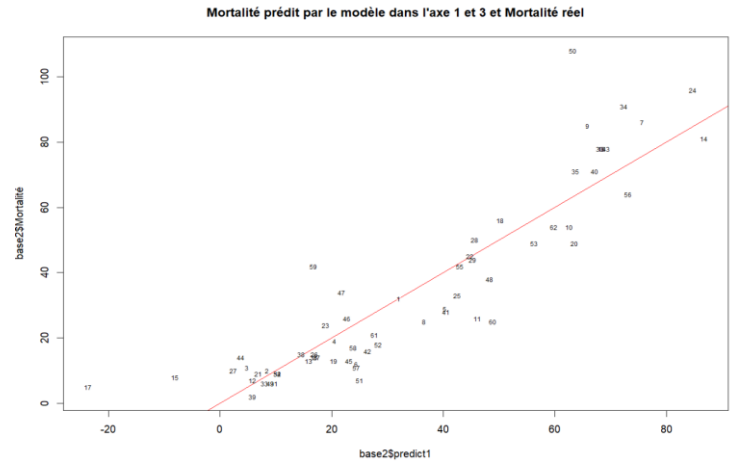


Figure n° 43 : Modèle prédit



Maintenant que nous avons effectué la régression linéaire multiple et que nous savons que les axes 1 et 3 ont un impact significatif sur la mortalité des enfants de moins de cinq ans, nous pouvons alors comparer les données prédites par le modèle et les données observées. Nous avons donc comparé les données prédites avec dans un premier temps le modèle prenant en compte l'axe 1 et 2 puis avec le modèle prenant en compte l'axe 1 et 3. Nous pouvons voir que les figures n° 42 et 43 sont assez proches l'une de l'autre. En effet, les valeurs de la mortalité infanto-juvénile prédites sont globalement assez proches de la droite rouge dans les deux cas. La droite rouge correspond à une situation parfaite, c'est-à-dire que les valeurs prédites sont égales aux valeurs observées. Les valeurs prédites par le modèle sont donc proches de la valeur observée. La proximité de ses points prédits par le modèle avec la droite rouge signifie donc une bonne prédiction de la part du modèle comme pour les points 22 et 43 sur la figure n°42 ou les points 22 et 38 sur la figure n°43. Cependant, certains points sont assez écartés de la ligne rouge comme les points 17 et 50 sur les deux nuages de points.

C. Régression logistique binaire

Nous allons maintenant faire une régression logistique binaire c'est-à-dire que nous allons transformer notre variable illustrative en variable binaire. Pour cela, nous allons diviser notre variable en deux groupes : le premier aura la valeur 1 et le deuxième la valeur 0. La valeur 1 correspond au succès de notre test c'est-à-dire, que les pays qui se

trouvent dans ce groupe ont une mortalité infanto-juvénile supérieure au seuil donné. Nous pouvons diviser la variable en deux groupes grâce à la valeur seuil, pour cela nous avons décidé d'utiliser la médiane qui correspond à 25. Dans ce cas, l'axe 1 est significatif au seuil de 5% (p-value inférieure à 0,05) mais les axes 2 et 3 ne sont pas significatifs.

Figure n° 44 : Logit au seuil de 25 axe1 et 2

```
Call:
glm(formula = base2$MortalitéN ~ res.pca$ind$coord[, 1] + res.pca$ind$coord[, 2], family = binomial)

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)    0.36184    0.54717   0.661  0.50842
res.pca$ind$coord[, 1] -1.74481    0.46398  -3.760  0.00017 ***
res.pca$ind$coord[, 2] -0.08659    0.39481  -0.219  0.82640
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 85.886  on 61  degrees of freedom
Residual deviance: 29.289  on 59  degrees of freedom
AIC: 35.289

Number of Fisher Scoring iterations: 7
```

Figure n° 45 : Logit au seuil de 25axe1 et 3

```
Call:
glm(formula = base2$MortalitéN ~ res.pca$ind$coord[, 1] + res.pca$ind$coord[, 3], family = binomial)

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)    0.32616    0.52557   0.621   0.535
res.pca$ind$coord[, 1] -1.70759    0.42366  -4.031 5.56e-05 ***
res.pca$ind$coord[, 3] -0.05551    0.52678  -0.105   0.916
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 85.886  on 61  degrees of freedom
Residual deviance: 29.327  on 59  degrees of freedom
AIC: 35.327

Number of Fisher Scoring iterations: 6
```

Une fois la régression faite avec la médiane, nous avons voulu avoir le seuil le plus optimal possible, c'est-à-dire avec un AIC le plus faible possible et aussi des variables latentes significatives si possible. L'AIC correspondant au critère d'information d'Akaike, doit être le plus faible possible car cela signifie que l'on cherche un modèle qui explique efficacement les données tout en restant le plus simple possible. Pour cela, nous avons ensuite essayé de faire la régression avec le quartile 1 de 12,25 (Annexe n°10 et 11). Cependant, l'AIC était supérieur comparé à la régression précédente mais les variables latentes étaient toutes significatives alors qu'avec le premier cas seule la variable latente de l'axe 1 était significative au seuil de risque de 0,05.

Nous avons donc cherché une valeur comprise entre le quartile 1 et la médiane [12,25 ;25].A l'aide de quelques tests, nous avons choisi la valeur de 20, correspondant à 20 enfants décédés avant l'âge de cinq ans pour 1 000 naissances vivantes. En effet, l'AIC est plus faible pour l'axe1 et 3 en comparaison avec la régression logistique binaire précédente. De plus, la variable latente de l'axe 3 devient elle aussi significative au seuil de risque de 10% et l'axe 1 reste significatif au seuil de 5%. Pour l'axe 1 et 2, l'AIC est légèrement supérieur mais seul l'axe 1 est significatif comme pour les tests précédents. L'axe 1 reste significatif au seuil de 5% tandis que l'axe2 reste non significatif. Les résultats de la régression sont présentés ci-dessous.

Figure n° 46 : Logit au seuil de 20 axes 1 et 2

```
Call:
glm(formula = base2$MortalitéN ~ res.pca$ind$coord[, 1] + res.pca$ind$coord[,
  2], family = binomial)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)      1.5394    0.7574   2.032 0.042104 *
res.pca$ind$coord[, 1] -2.0705    0.5857  -3.535 0.000408 ***
res.pca$ind$coord[, 2] -0.2244    0.4072  -0.551 0.581547
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 85.369  on 61  degrees of freedom
Residual deviance: 29.301  on 59  degrees of freedom
AIC: 35.301

Number of Fisher Scoring iterations: 7
```

Figure n° 47 : Logit au seuil de 20 axes 1 et 3

```
Call:
glm(formula = base2$MortalitéN ~ res.pca$ind$coord[, 1] + res.pca$ind$coord[,
  3], family = binomial)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)      1.6894    0.8678   1.947 0.05157 .
res.pca$ind$coord[, 1] -2.3556    0.7356  -3.202 0.00136 **
res.pca$ind$coord[, 3] -1.1326    0.6358  -1.781 0.07484 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 85.369  on 61  degrees of freedom
Residual deviance: 25.780  on 59  degrees of freedom
AIC: 31.78

Number of Fisher Scoring iterations: 8
```

Nous allons maintenant interpréter la régression logistique binaire avec un seuil de 20. Tout d'abord, dans le plan factoriel des axes 1 et 2, l'axe 1 a un impact significatif sur la mortalité au seuil de risque de 5% et l'axe 2 n'a pas d'impact significatif. Tandis que sur le plan factoriel des axes 1 et 3, l'axe 1 a toujours un impact significatif. L'axe 3 a quant à lui, un impact significatif sur la mortalité mais au seuil de risque de 10%. Cependant, pour connaître le réel impact de l'axe 1 et 3, nous devons calculer l'exponentiel du coefficient donné par « estimate std ». Nous pouvons dire que dans le premier cas, la probabilité pour un pays d'être dans le groupe où la mortalité est supérieure à 20 décès pour 1 000 naissances vivantes sera 0,126 ($=\exp(-2,0705)$) fois inférieur lorsque l'on se déplace vers la droite de l'axe 1.

Dans le deuxième plan, la probabilité d'être dans le groupe dont la mortalité est supérieure à 20 sera 0,0948($=\exp(-2,3556)$) fois inférieur lorsque l'on se déplace vers la droite de l'axe 1. De plus, la probabilité d'être dans ce même groupe sera 0,323 ($=\exp(-1,1326)$) fois supérieure lorsque l'on se déplace vers le haut de l'axe 3. En effet, lorsque l'on se déplace vers la droite de l'axe 1, la situation sanitaire s'améliore et par conséquent, la probabilité d'avoir un taux de mortalité infanto-juvénile élevé diminue. Pour l'axe 3, en se déplaçant vers le haut de l'axe 3, le niveau de scolarisation diminue, cela explique pourquoi la probabilité d'être dans le groupe avec une mortalité supérieur à 20 décès pour 10 000 naissances vivantes augmente.

V. Conclusion et discussion des résultats

Nous sommes parties de la littérature pour construire un modèle pouvant expliquer la mortalité infanto-juvénile. Nous pouvons dire que les résultats des différents tests sont satisfaisants et cohérents avec la littérature. Nous avons dans un premier temps, réalisé une étude des corrélations qui nous a permis de visualiser à l'aide de différents graphiques, la relation entre les variables actives et la variable illustrative.

Par la suite, l'analyse statistique a fourni des informations sur la distribution de l'ensemble des données. Cela nous a permis d'observer la véracité de nos données.

Ensuite, nous avons poursuivi notre étude en effectuant une analyse en composantes principales. Cette dernière, nous a permis d'identifier trois variables latentes synthétisant au mieux les douze variables explicatives. A l'issue de cette ACP, nous avons pu constater que les pays ayant à la fois une mauvaise situation sanitaire et/ou un faible taux de scolarisation et/ou un progrès social faible sont plus enclin à avoir un fort taux de mortalité des enfants de moins de cinq ans.

Enfin, nous avons terminé notre analyse par une régression linéaire multiple. Celle-ci nous a permis de constater d'une part, la non-significativité de l'axe 2 (progrès social) et d'autre part la significativité de l'axe 1 (situation sanitaire) et 3 (scolarisation). Cela signifie que les variables latentes représentant l'axe 1 et 3 expliquent bien le taux de mortalité infanto-juvénile. Par ailleurs, la régression linéaire binaire permet de séparer notre groupe d'individus en deux selon leur taux de mortalité conformément au seuil de 20 décès pour 1 000 naissances vivantes. Dans cette analyse, l'axe 1 et 3 sont toujours significatifs alors que l'axe 2 ne l'est toujours pas.

Toutefois, certaines données auraient pu être intéressantes d'étudier pour compléter notre analyse. Par exemple, la pollution de l'air ou la consommation de substances néfastes pour la santé telles que l'alcool auraient pu jouer un rôle sur le taux de mortalité. Cependant, nous n'avons pas réussi à trouver de telles données. De plus, nous n'avons pas inclus la guerre comme facteur, alors qu'il engendre un grand nombre de décès chez les enfants comme nous avons pu le voir dans les actualités.

VI. Annexes

A. Tableaux et graphiques supplémentaires

Annexe n° 1 : Comparaison base avec et sans les valeurs atypiques

```
> #Comparaison base avec et sans valeurs atypiques
> skewness(base2$Mortalité)
[1] 0.919495
> kurtosis(base2$Mortalité)
[1] 2.694601
> skewness(BaseTypique2$Mortalité)
[1] 1.120309
> kurtosis(BaseTypique2$Mortalité)
[1] 3.41148
```

Annexe n° 2 : Caractéristiques des individus

Coordonnées				Contributions				Cosinus2			
Individus	Dim.1	Dim.2	Dim.3	Individus	Dim.1	Dim.2	Dim.3	Individus	Dim.1	Dim.2	Dim.3
1	0.26	-1.26	-0.14	1	0,02	1,89	0,03	1	0.01	0.24	0.00
2	2.20	0.21	0.44	2	1,5	0,05	0,27	2	0.65	0.01	0.03
3	2.15	-0.62	1.34	3	1,43	0,45	2,49	3	0.51	0.04	0.20
4	0.82	-0.71	1.05	4	0,21	0,59	1,55	4	0.09	0.06	0.14
5	-0.49	2.13	-0.15	5	0,08	5,39	0,03	5	0.02	0.43	0.00
6	1.23	-0.58	-0.74	6	0,47	0,39	0,76	6	0.42	0.09	0.15
7	-3.47	-0.13	-0.86	7	3,74	0,02	1,02	7	0.84	0.00	0.05
8	0.14	-1.50	-0.82	8	0,01	2,66	0,93	8	0.00	0.36	0.11
9	-2.96	0.12	0.06	9	2,72	0,02	0,01	9	0.59	0.00	0.00
10	-2.07	0.84	-1.28	10	1,32	0,85	2,29	10	0.48	0.08	0.18
11	-1.35	1.43	0.58	11	0,57	2,43	0,46	11	0.31	0.34	0.06
12	2.73	0.29	-0.21	12	2,31	0,1	0,06	12	0.62	0.01	0.00
13	2.23	-0.41	-1.22	13	1,54	0,2	2,08	13	0.51	0.02	0.15
14	-4.01	-0.85	-1.97	14	4,99	0,87	5,4	14	0.63	0.03	0.15
15	4.59	-0.25	-1.51	15	6,52	0,07	3,19	15	0.70	0.00	0.08
16	-3.10	-0.35	-0.16	16	2,98	0,14	0,03	16	0.84	0.01	0.00
17	5.00	-0.89	0.87	17	7,74	0,94	1,04	17	0.68	0.02	0.02
18	-2.81	-0.03	3.05	18	2,44	0	12,95	18	0.29	0.00	0.35
19	1.45	-1.33	-0.36	19	0,65	2,12	0,18	19	0.36	0.31	0.02
20	-2.38	-0.78	-0.78	20	1,75	0,73	0,85	20	0.53	0.06	0.06
21	2.24	-0.72	0.69	21	1,56	0,61	0,67	21	0.49	0.05	0.05
22	-1.07	0.42	0.21	22	0,36	0,21	0,06	22	0.30	0.05	0.01
23	1.82	-1.08	-0.92	23	1,03	1,38	1,19	23	0.40	0.14	0.10
24	-4.69	-2.25	-0.01	24	6,79	6,04	0	24	0.67	0.16	0.00
25	-1.70	1.35	2.13	25	0,89	2,16	6,33	25	0.19	0.12	0.31
26	1.01	-1.00	1.38	26	0,32	1,19	2,66	26	0.15	0.14	0.27
27	2.65	-0.61	0.72	27	2,17	0,44	0,73	27	0.60	0.03	0.04
28	-0.89	2.41	-0.38	28	0,25	6,93	0,2	28	0.05	0.35	0.01
29	-1.33	0.33	0.71	29	0,55	0,13	0,69	29	0.51	0.03	0.14
30	-3.43	-1.94	0.67	30	3,65	4,46	0,63	30	0.59	0.19	0.02
31	2.12	-0.83	0.38	31	1,39	0,82	0,2	31	0.81	0.12	0.03
32	1.84	1.00	0.90	32	1,04	1,19	1,13	32	0.35	0.10	0.08
33	1.76	3.49	1.56	33	0,95	14,51	3,4	33	0.09	0.35	0.07
34	-4.21	-1.30	1.54	34	5,47	2,01	3,3	34	0.71	0.07	0.09
35	-2.95	-0.28	0.49	35	2,7	0,1	0,34	35	0.84	0.01	0.02
36	1.95	-1.16	-0.76	36	1,17	1,59	0,8	36	0.48	0.17	0.07
37	1.65	-0.55	-0.17	37	0,84	0,35	0,04	37	0.71	0.08	0.01
38	1.83	-0.23	-0.01	38	1,04	0,06	0	38	0.46	0.01	0.00
39	2.65	-0.81	-0.00	39	2,17	0,78	0	39	0.85	0.08	0.00
40	-2.51	0.48	-1.25	40	1,96	0,28	2,17	40	0.59	0.02	0.15
41	-0.60	1.43	0.03	41	0,11	2,44	0	41	0.04	0.25	0.00
42	1.33	-0.13	-1.39	42	0,55	0,02	2,71	42	0.21	0.00	0.23
43	-3.88	-1.34	1.46	43	4,65	2,13	2,96	43	0.60	0.07	0.09
44	2.33	-0.47	1.17	44	1,68	0,26	1,9	44	0.29	0.01	0.07
45	1.82	-0.76	-1.80	45	1,03	0,68	4,52	45	0.37	0.06	0.37
46	0.87	-1.11	0.46	46	0,23	1,46	0,3	46	0.12	0.19	0.03
47	1.47	-0.48	-0.72	47	0,67	0,27	0,72	47	0.43	0.05	0.10
48	-1.82	0.65	1.19	48	1,02	0,5	1,99	48	0.43	0.05	0.18
49	1.98	-0.48	0.84	49	1,21	0,28	0,97	49	0.52	0.03	0.09
50	-1.43	2.01	-2.90	50	0,63	4,81	11,75	50	0.08	0.15	0.32
51	0.14	1.22	1.65	51	0,01	1,76	3,81	51	0.00	0.17	0.32
52	0.80	-1.33	-0.58	52	0,2	2,11	0,47	52	0.06	0.16	0.03
53	-1.78	-0.07	-0.61	53	0,98	0,01	0,52	53	0.48	0.00	0.06
54	1.95	0.51	0.64	54	1,18	0,31	0,58	54	0.48	0.03	0.05
55	-0.94	1.62	0.30	55	0,28	3,1	0,13	55	0.16	0.48	0.02
56	-2.93	0.55	-1.56	56	2,66	0,36	3,41	56	0.65	0.02	0.19
57	0.97	2.88	-0.14	57	0,29	9,87	0,03	57	0.05	0.41	0.00
58	1.20	0.57	-0.54	58	0,44	0,39	0,4	58	0.38	0.09	0.08
59	1.77	0.70	-0.34	59	0,97	0,58	0,16	59	0.50	0.08	0.02
60	-0.92	1.41	-1.01	60	0,26	2,35	1,41	60	0.10	0.24	0.12
61	0.92	0.91	-0.73	61	0,26	0,99	0,74	61	0.16	0.16	0.10
62	-2.14	-0.39	-0.52	62	1,42	0,18	0,38	62	0.54	0.02	0.03

Annexe n° 3 : Tableau affichant les contributions des pays selon les axes

CORRELATION	POSITIVE	CONTRIBUTION	NEGATIVE	CONTRIBUTION
AXE 1	17 : Cuba	7,74	24 : Guinée	6,79
	15 : Costa Rica	6,52	34 : Mali	5,47
	12 : Chine	2,31	14 : Congo (démocratique)	5,47
	27 : Kazakhstan	2,17	43 : Niger	4,65
	39 : Monténégro	2,17	7 : Bénin	3,74
	44 : Ouzbékistan	1,68	30 : Libéria	3,65
			16 : Côte d'Ivoire	2,98
			9 : Burkina Faso	2,72
AXE2	33 : Maldives	14,51	24 : Guinée	6,04
	57 : Tonga	9,87	30 : Libéria	4,46
	28 : Kiribati	6,93	8 : Bolivie	2,66
	5 : Bangladesh	5,39	43 : Niger	2,13
	50 : Sierra Leone	4,81	19 : Equateur	2,12
	55 : Timor-Leste	3,10	52 : Suriname	2,11
	41 : Népal	2,44	34 : Mali	2,01
AXE 3	18 : Djibouti	12,95	50 : Sierra Leone	11,75
	25 : Inde	6,33	14 : Congo (démocratique)	5,40
	51 : Sierra Leone	3,81	45 : Pérou	4,52
	33 : Maldives	3,40	56 : Togo	3,41
	34 : Mali	3,30	15 : Costa Rica	3,19
	43 : Niger	2,71	10 : Burundi	2,29

Annexe n° 4 : Test Shapiro axes 1 et 3

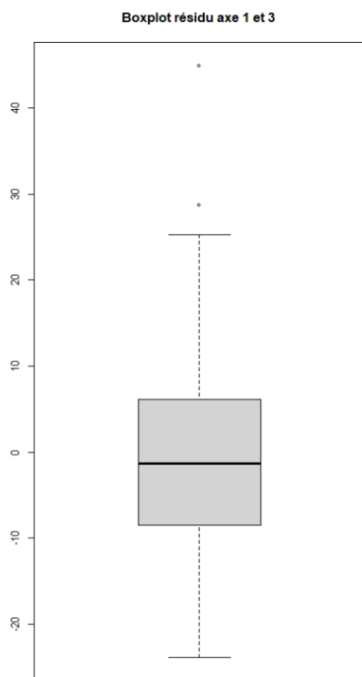
```
> shapiro.test(residuals(modele))
```

Shapiro-wilk normality test

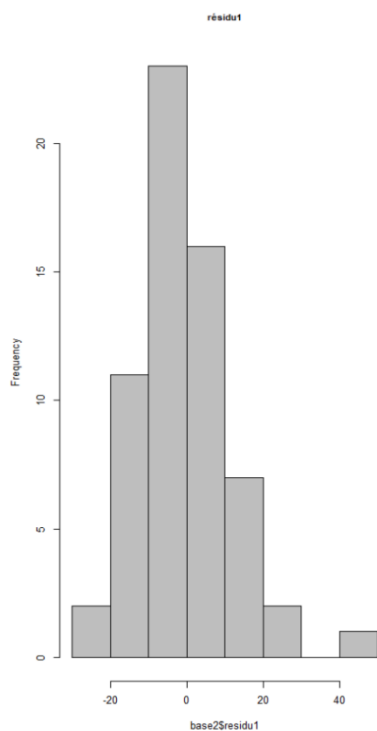
```
data: residuals(modele)
```

```
W = 0.84855, p-value = 2.02e-06
```

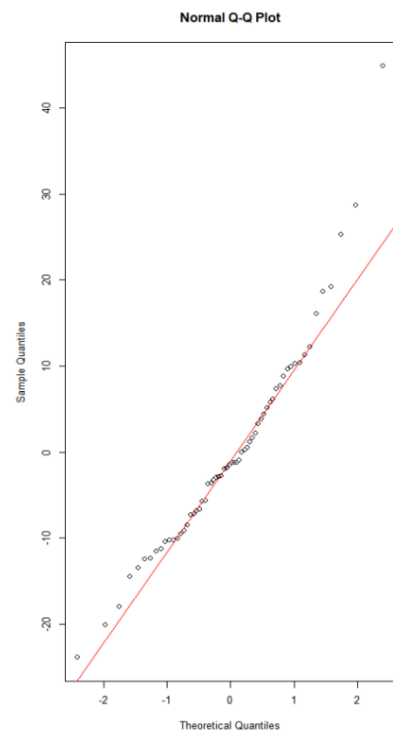
Annexe n° 5 : Boxplot des résidus : axes 1 et 3



Annexe n° 6 : histogramme de résidus axes 1 et 3



Annexe n° 7 : Q-Q plot des : résidus axes 1 et 3



Annexe n° 8 : Test Kolmogorov-Smirnow axes 1 et 3

```
> ks.test(base2$residu1,"pnorm",mean(base2$residu1),sd(base2$residu1))
```

Asymptotic one-sample Kolmogorov-Smirnov test

```
data: base2$residu1
D = 0.097139, p-value = 0.6022
alternative hypothesis: two-sided
```

Annexe n° 9 : Test Breusch-Pagan axes 1 et 3

```
> bptest(modele1)
```

studentized Breusch-Pagan test est

```
data: modele1
BP = 5.4346, df = 2, p-value = 0.06605 605
```

Annexe n° 10 : Test Logit au seuil de 12,25, axes 1 et 2

```
Call:
glm(formula = base2$MortalitéN ~ res.pca$ind$coord[, 1] + res.pca$ind$coord[, 2], family = binomial)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	4.0923	1.3348	3.066	0.00217 **
res.pca\$ind\$coord[, 1]	-2.3342	0.7622	-3.063	0.00219 **
res.pca\$ind\$coord[, 2]	-1.3728	0.6316	-2.173	0.02975 *

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 70.807 on 61 degrees of freedom
Residual deviance: 31.498 on 59 degrees of freedom
AIC: 37.498

Number of Fisher Scoring iterations: 8

Annexe n° 11 : Test Logit au seuil de 12,25, axes 1 et 3

```
Call:
glm(formula = base2$MortalitéN ~ res.pca$ind$coord[, 1] + res.pca$ind$coord[, 3], family = binomial)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	3.4627	1.0731	3.227	0.00125 **
res.pca\$ind\$coord[, 1]	-1.7947	0.5618	-3.195	0.00140 **
res.pca\$ind\$coord[, 3]	-1.5071	0.5784	-2.605	0.00918 **

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 70.807 on 61 degrees of freedom
Residual deviance: 30.669 on 59 degrees of freedom
AIC: 36.669

Number of Fisher Scoring iterations: 7

B. Base de données

▲	Individus	Code pays	Mortalité	Médecins	Maigreur extrême	Infirmiers et personnels obstétriques	Dépenses publiques	Hépatite B	Vaccin DTC3	PIB/hab	Anémie	Recherche et santé	Scolarisation	Eau	Espérance
1	Afrique du Sud	ZAF	32	7.9	3.4	49.7	15.3	3.10	84	5741.6412	30.5	0.81	95.37994	93.76479	65.25200
2	Albanie	ALB	10	18.8	1.6	60.5	9.0	0.29	98	5343.0377	24.8	2.03	101.58872	95.06804	76.98900
3	Arménie	ARM	11	44.0	4.4	49.5	7.7	0.06	91	4505.8677	17.3	3.21	91.77408	99.94152	72.17300
4	Azerbaïdjan	AZE	19	31.7	3.2	64.3	4.0	0.06	79	4229.9106	35.1	0.59	98.03700	96.36208	66.86800
5	Bangladesh	BGD	29	6.7	9.8	4.9	3.1	0.51	98	2233.3059	36.7	2.81	124.98738	97.94568	71.96800
6	Belize	BLZ	12	10.8	1.8	23.4	12.6	0.60	79	5266.8762	20.5	3.71	105.15975	98.40196	72.85400
7	Bénin	BEN	86	0.6	5.0	3.0	4.6	2.87	72	1237.9493	55.2	5.80	109.67425	66.88733	60.08800
8	Bolivie	BOL	25	10.3	2.0	15.6	14.9	0.14	68	3068.8126	24.4	2.28	98.76400	93.08742	64.46700
9	Burkina Faso	BFA	85	0.9	8.1	9.3	11.5	1.66	91	833.2443	52.5	6.34	93.05973	48.54099	59.73100
10	Burundi	BDI	54	0.7	4.8	6.5	8.3	1.35	93	216.8274	38.5	8.09	115.88058	62.20712	61.56600
11	Cambodge	KHM	26	1.9	9.7	10.1	7.4	0.19	92	1577.9117	47.1	10.71	103.70227	76.25311	70.41600
12	Chine	CHN	7	22.3	1.9	30.8	8.4	0.22	99	10408.7191	15.5	0.06	103.71258	95.89031	78.07700
13	Colombie	COL	13	23.3	1.6	14.6	19.5	0.15	88	5304.2891	21.2	0.40	110.71261	97.04102	74.76900
14	Congo (démocratique)	COD	81	3.8	6.4	11.1	6.3	3.28	57	524.6667	42.4	6.70	120.54784	35.50024	59.73900
15	Costa Rica	CRI	8	33.0	1.8	38.0	25.2	0.02	95	12179.2567	13.7	0.95	115.99692	99.81054	79.27700
16	Côte d'Ivoire	CIV	78	1.6	6.1	6.6	6.7	3.44	80	2349.0699	50.9	6.51	97.53019	72.62534	59.03200
17	Cuba	CUB	5	84.2	2.0	75.6	16.4	0.03	99	9499.5902	19.3	0.97	101.63053	94.26628	77.56700
18	Djibouti	DJI	56	2.2	21.5	7.3	4.3	0.26	70	2917.9963	32.3	14.02	72.20651	76.04992	62.69400
19	Equateur	ECU	13	22.2	3.7	25.1	13.9	0.09	70	5645.1993	17.2	0.67	100.08103	95.35976	72.15300
20	Ethiopie	ETH	49	1.1	7.2	7.8	6.8	1.59	71	918.6526	23.9	3.17	102.33886	48.65535	65.37100
21	Géorgie	GEO	9	51.1	0.6	55.5	8.1	0.06	88	4255.7430	27.5	1.99	99.67834	94.67210	72.76500
22	Ghana	GHA	45	1.7	6.8	36.2	6.9	2.13	94	2176.5762	35.4	4.09	104.36254	86.31681	64.11400
23	Guatemala	GTM	24	12.4	0.8	22.4	15.9	0.03	83	4609.8973	7.4	1.45	99.65825	93.88009	71.79700
24	Guinée	GIN	96	2.2	9.2	5.8	6.2	6.07	47	1073.6593	48.0	5.40	92.44959	68.92063	59.32700
25	Inde	IND	33	7.4	17.3	17.5	3.3	0.16	85	1913.2197	53.0	0.26	100.90824	92.14560	70.15000
▲	Individus	Code pays	Mortalité	Médecins	Maigreur extrême	Infirmiers et personnels obstétriques	Dépenses publiques	Hépatite B	Vaccin DTC3	PIB/hab	Anémie	Recherche et santé	Scolarisation	Eau	Espérance
26	Jordanie	JOR	15	26.6	2.4	33.5	11.9	0.56	77	3987.6502	37.7	11.29	79.83081	98.96051	75.18400
27	Kazakhstan	KAZ	10	40.7	3.1	72.9	10.2	0.15	88	9121.6364	28.7	0.39	100.20895	95.43497	71.37000
28	Kiribati	KIR	50	2.0	3.5	38.3	8.4	1.57	92	1403.9937	32.6	58.92	111.12331	74.30948	67.26600
29	Laos	LAO	44	3.5	9.0	11.9	6.2	0.68	79	2593.3551	39.5	8.61	97.16892	85.21974	68.49700
30	Libéria	LBR	78	0.5	3.4	19.5	4.5	4.66	65	597.5297	42.6	10.02	77.27743	74.65215	60.94800
31	Macédoine du Nord	MKD	6	28.7	3.4	37.9	12.8	0.08	84	5965.4502	19.3	1.42	95.45911	97.78244	74.39512
32	Malaisie	MYS	9	22.9	9.7	34.8	8.6	0.06	98	10160.8293	32.0	0.17	104.63542	97.09991	75.93800
33	Maldives	MDV	6	20.5	9.1	46.6	18.2	0.21	99	7282.3729	52.2	69.47	100.63843	99.54445	79.87500
34	Mali	MLI	91	1.3	9.3	4.4	5.7	4.62	70	822.9061	59.0	6.74	74.81614	80.74719	58.63300
35	Mauritanie	MRT	71	1.9	11.5	9.3	7.0	3.35	71	1868.4682	43.3	9.77	97.63359	74.29672	64.53200
36	Mexique	MEX	14	24.3	1.4	28.2	11.5	0.03	74	8654.9973	15.3	0.07	104.18820	99.36870	70.13300
37	Moldavie	MDA	14	31.0	1.9	46.8	12.4	0.20	86	4376.2425	26.1	2.92	105.27157	91.13725	70.16600
38	Mongolie	MNG	15	38.5	0.9	42.1	8.5	0.47	96	4041.1741	14.5	7.93	101.73135	82.12412	72.14100
39	Monténégro	MNE	2	27.4	2.2	53.7	13.1	0.67	84	7677.3691	17.2	1.64	101.63207	98.85691	75.93171
40	Mozambique	MOZ	71	0.8	4.4	4.8	7.3	0.59	79	454.0624	47.9	8.09	114.92888	59.76011	61.17200
41	Népal	NPL	28	8.5	12.0	33.4	5.7	0.16	84	1139.1899	35.7	4.03	126.26804	90.76858	69.24600
42	Nicaragua	NIC	16	16.6	2.2	15.5	18.4	0.09	92	1876.6074	15.7	4.83	111.78294	81.70875	71.79500
43	Niger	NER	78	0.3	9.8	2.2	10.2	3.44	81	564.8220	49.5	5.43	65.59661	47.72058	61.45100
44	Ouzbékistan	UZB	14	23.7	1.8	112.8	10.8	0.16	95	1759.3075	24.8	3.03	99.11991	95.95781	70.33100
45	Pérou	PER	13	13.7	0.4	29.8	16.3	0.06	72	6063.6266	20.6	0.72	118.67553	93.60797	73.66500
46	Philippines	PHL	26	7.7	5.6	54.4	8.7	0.38	71	3224.4228	12.3	1.87	97.05741	94.05123	72.11900
47	République dominicaine	DOM	34	14.5	2.4	14.6	14.7	0.10	82	7167.9149	26.4	0.84	103.33073	96.30658	72.88900
48	Senegal	SEN	38	0.9	8.1	5.4	6.5	0.93	91	1492.4700	52.7	6.13	88.91621	83.88690	68.00600
49	Serbie	SRB	6	31.1	2.6	60.9	11.0	0.04	92	1492.4759	22.8	1.37	97.64904	95.29553	74.47805
50	Sierra Leone	SLE	108	0.7	5.4	7.5	5.3	1.98	91	12020.0219	48.4	8.53	143.67078	63.06756	59.76300
▲	Individus	Code pays	Mortalité	Médecins	Maigreur extrême	Infirmiers et personnels obstétriques	Dépenses publiques	Hépatite B	Vaccin DTC3	PIB/hab	Anémie	Recherche et santé	Scolarisation	Eau	Espérance
51	Sri Lanka	LKA	7	12.3	15.1	25.0	8.5	0.34	96	3852.3900	34.6	0.86	100.33962	88.69595	76.39300
52	Suriname	SUR	18	8.2	5.5	39.3	13.8	0.07	51	4796.5333	21.0	7.41	108.48266	97.98963	72.56100
53	Tanzanie	TZA	49	0.5	3.5	5.7	9.4	0.99	86	1104.1644	38.9	5.56	98.36622	57.97027	66.40800
54	Thaïlande	THA	9	9.5	7.7	31.5	13.2	0.27	97	7001.7854	24.0	0.24	98.96024	100.00000	79.27400
55	Timor-Leste	TLS	42	7.6	9.9	17.5	6.6	0.72	86	1660.3083	29.9	29.44	109.64244	83.97422	68.47500
56	Togo	TGO	64	0.8	5.7	5.1	5.4	3.27	82	875.2454	45.7	4.67	124.42168	68.89106	61.03500
57	Tonga	TON	11	9.5	1.1	45.4	7.9	0.89	99	4605.9709	28.5	62.37	112.11738	98.73746	70.92800
58	Tunisie	TUN	17	13.0	2.1	25.1	10.8	0.12	92	3497.7332	32.1	2.65	112.53810	96.69266	75.29200
59	Turkménistan	TKM	42	22.2	4.1	44.3	8.7	0.07	98	7297.1800	26.6	0.66	115.91598	100.00000	68.68700
60	Vanuatu	VUT	25	1.7	4.7	14.2	5.9	2.25	78	2917.7568	28.5	33.94	120.43426	90.48903	70.29900
61	Vietnam	VNM	21	8.3	5.8	14.5	9.4	0.64	94	3586.3473	20.6	1.26	118.56549	96.66702	75.37800
62	Zimbabwe	ZWE	54	2.0	2.9	21.4	5.2	2.74	86	1372.6967	28.9	10.96	97.55364	62.66646	61.12400

C. Script R

#Les librairies utilisées

```
library(readxl) #Lire le excel
library(outliers)
library(EnvStats)
library(FactoMineR)#Faire l'ACP
library(moments)
library(factoextra)
library(gridExtra) #faire tableau grid-table
library(RColorBrewer)
library(corrplot) #Faire les corrplot
library(lmtest)
library(forecast)
library(gtsummary)
```

#Renommer les variables

```
colnames(Donnée)
names(Donnée)[names(Donnée) == "Dépenses publiques\r\n\r\n"] <- "Dépenses
publiques"
names(Donnée)[names(Donnée) == "Infirmiers et personnels obstétriques\r\n\r\n"] <-
"Infirmiers et personnels obstétriques"
names(Donnée)[names(Donnée) == "Espérance de vie à la naissance\r\n"] <-
"Espérance"
names(Donnée)[names(Donnée) == "Eau\r\n"] <- "Eau"
colnames(Donnée)
View(Donnée)
```

#Nettoyer la base

```
BaseA<-na.omit(Donnée)
```

#Base sans la variable illustrative

```
base1<-BaseA[,4:15]
```

#Base avec la variable illustrative

```
base2<-BaseA[,3:15]
```

#Analyse univariée sous forme de tableau

```
summary(base2)
```

```
tbl_summary(base2, statistic = all_continuous() ~ "Moy = {mean} [Etendue = {max}-  
{min}] écart-type = {sd} Q1 = {p25} médiane={median} et Q3= {p75}")
```

```
quantile(base2)
```

#Graphiques de corrélation entre la variable illustrative et les variables actives

```
plot(BaseA$`Prévalence Hépatite B`, BaseA$Mortalité, type="p", main="Corrélation  
Mortalité et Hépatite B", xlab="Hépatite B en %", ylab="Mortalité infanto-juvénile",  
font.lab=3) abline(lm(BaseA$Mortalité~BaseA$`Prévalence Hépatite  
B`), col="red", lwd=2)
```

```
plot(BaseA$`Anémie`, BaseA$Mortalité, type="p", main="Corrélation Mortalité et  
Anémie", xlab="Anémie en %", ylab="Mortalité infanto-juvénile", font.lab=3)
```

```
abline(lm(BaseA$Mortalité~BaseA$`Anémie`), col="red", lwd=2)
```

```
plot(BaseA$`Maigreur extrême`, BaseA$Mortalité, type="p", main="Corrélation Mortalité  
et Maigreur extrême", xlab="Maigreur extrême en %", ylab="Mortalité infanto-juvénile",  
font.lab=3)
```

```
abline(lm(BaseA$Mortalité~BaseA$`Maigreur extrême`), col="red", lwd=2)
```

```
plot(BaseA$`Recherche et santé`, BaseA$Mortalité, type="p", main="Corrélation Mortalité  
et Recherche et santé", xlab="Recherche et santé", ylab="Mortalité infanto-juvénile",  
font.lab=3)
```

```
abline(lm(BaseA$Mortalité~BaseA$`Recherche et santé`), col="red", lwd=2)
```

```
plot(BaseA$`Scolarisation`, BaseA$Mortalité, type="p", main="Corrélation Mortalité et  
Scolarisation", xlab="Scolarisation", ylab="Mortalité infanto-juvénile", font.lab=3)
```

```
abline(lm(BaseA$Mortalité~BaseA$`Scolarisation`), col="red", lwd=2)
```

```
plot(BaseA$`Vaccin DTC3`, BaseA$Mortalité, type="p", main="Corrélation Mortalité et  
Vaccin DTC3", xlab="Vaccin DTC3", ylab="Mortalité infanto-juvénile", font.lab=3)
```

```
abline(lm(BaseA$Mortalité~BaseA$`Vaccin DTC3`), col="red", lwd=2)
```

```
plot(BaseA$`PIB/hab`,BaseA$Mortalité,type="p", main="Corrélation Mortalité et PIB/hab",xlab="PIB/hab",ylab="Mortalité infanto-juvénile", font.lab=3)
```

```
abline(lm(BaseA$Mortalité~BaseA$`PIB/hab`),col="red",lwd=2)
```

```
plot(BaseA$`Eau`,BaseA$Mortalité,type="p", main="Corrélation Mortalité et Eau",xlab="Eau",ylab="Mortalité infanto-juvénile", font.lab=3)
```

```
abline(lm(BaseA$Mortalité~BaseA$`Eau`),col="red",lwd=2)
```

```
plot(BaseA$`Médecins`,BaseA$Mortalité,type="p", main="Corrélation Mortalité et Médecins",xlab="Médecins",ylab="Mortalité infanto-juvénile", font.lab=3)
```

```
abline(lm(BaseA$Mortalité~BaseA$`Médecins`),col="red",lwd=2)
```

```
plot(BaseA$`Infirmière et sage-femme`,BaseA$Mortalité,type="p", main="Corrélation Mortalité et Infirmière et sage-femme",xlab="Infirmière et sage-femme",ylab="Mortalité infanto-juvénile", font.lab=3)
```

```
abline(lm(BaseA$Mortalité~BaseA$`Infirmière et sage-femme`),col="red",lwd=2)
```

```
plot(BaseA$`Dépenses publiques`,BaseA$Mortalité,type="p", main="Corrélation Mortalité et Dépenses publiques",xlab="Dépenses publiques",ylab="Mortalité infanto-juvénile", font.lab=3)
```

```
abline(lm(BaseA$Mortalité~BaseA$`Dépenses publiques`),col="red",lwd=2)
```

```
plot(BaseA$`Espérance`,BaseA$Mortalité,type="p", main="Corrélation Mortalité et Espérance",xlab="Espérance",ylab="Mortalité infanto-juvénile", font.lab=3)
```

```
abline(lm(BaseA$Mortalité~BaseA$`Espérance`),col="red",lwd=2)
```

#Boxplot afin d'observer les valeurs atypiques

```
boxplot(base2$Mortalité, xlab ="Mortlité infanto-juvénile")
```

```
par(mfrow=c(3,4))
```

```
boxplot(base2$Médecins, xlab ="Médecins pour 10 000 hab")
```

```
boxplot(base2$`Maigreur extrême`, xlab ="Maigreur extême")
```

```
boxplot(base2$`PIB/hab`, xlab ="PIB/hab")
```

```
boxplot(base2$Anémie, xlab ="Anémie")
```

```

boxplot(base2$Scolarisation, xlab = "Scolarisation")
boxplot(base2$Eau, xlab = "Eau")
boxplot(base2$`Recherche et santé`, xlab = "Recherche et santé")
boxplot(base2$`Vaccin DTC3`, xlab = "Vaccin DTC3 (%)")
boxplot(base2$`Prévalence Hépatite B`, xlab = "Prévalence Hepatite B (%)")
boxplot(base2$Espérance, xlab = "Espérance de vie")
boxplot(base2$`Infirmière et sage-femme`, xlab = "Infirmière et sage-femme")
boxplot(base2$`Dépenses publiques`, xlab = "Dépense publiques")

```

#Test de Grubbs pour vérifier la présence d'une valeur atypique

```

grubbs.test(base2$Mortalité,type=10,two.sided=TRUE)
#108 est atypique => sierra leone
grubbs.test(base2$Médecins,type=10,two.sided=TRUE)
#84,2 est atypique => Cuba
grubbs.test(base2$Eau,type=10,two.sided=TRUE)
#35,5 est atypique => Congo (démocratique)
grubbs.test(base2$`Infirmière et sage-femme`,type=10,two.sided=TRUE)
#112,8 est atypique => Ouzbékistan => vérifier sur les bases de données d'origine
grubbs.test(base2$`Dépenses publiques`,type=10,two.sided=TRUE)
#25,2 est atypique => Costa Rica => vérifier sur les bases de données d'origine

```

#Test de Rosner plus ouvert qui identifie jusqu'à dix valeurs atypiques

```

rosnerTest(base2$`Maigreur extrême`,k=2, alpha=0.05)
#21,5 et 17,3 sont atypiques => Djibouti et Inde
rosnerTest(base2$`PIB/hab`,k=2, alpha=0.05)
#Pas de valeur atypique
rosnerTest(base2$Scolarisation,k=5, alpha=0.05)
#Pas de valeur atypique
rosnerTest(base2$`Recherche et santé`,k=5, alpha=0.05)

```

69,47 et 62,37 et 58,92 et 33,94 et 29,44 sont atypiques => Maldives, Tonga, Kiribati, Vanuata et Timor-Leste

```
rosnerTest(base2$`Vaccin DTC3`,k=2, alpha=0.05)
```

#Pas de valeur atypique

```
rosnerTest(base2$`Prévalence Hépatite B`,k=4, alpha=0.05)
```

#6,07 est atypique => Guinée

#Enlever les pays ayant des valeurs atypiques

```
BaseTypique<- Donnée[-c(50,17,14,44,15,18,25,33,57,28,60,55,24),]
```

#Base ayant ni les valeurs atypiques ni la variable illustrative

```
BaseTypique1 <- BaseTypique[,4:15]
```

#Base ayant la variable illustrative sans les valeurs atypiques

```
BaseTypique2 <- BaseTypique[,3:15]
```

#Comparaison base avec et sans valeurs atypiques

```
skewness(base2$Mortalité)
```

```
kurtosis(base2$Mortalité)
```

```
skewness(BaseTypique2$Mortalité)
```

```
kurtosis(BaseTypique2$Mortalité)
```

#ACP

#Nous faisons l'ACP sur la base n'ayant pas la variable illustrative

```
res.pca=PCA(BaseA[,4:15])
```

#Valeurs propres

```
Valeurs_Propres<-round(res.pca$eig,2)
```

```
Valeurs_Propres
```

#Etude des variables

#Coordonnées des variables

```
Coordonnees_var<-round(res.pca$var$coord,2)
```



```
dim_coord<- Coordonnees_var[,1:3]

corrplot(dim_coord, is.corr=FALSE, method="circle",tl.srt=45, tl.col="#004400",
col=brewer.pal(n=9,name="Spectral"),addCoef.col="black")
```

#Contribution des variables

```
par(mfrow=c(1,1))

Contributions_var<-round(res.pca$var$contrib,2)

dim_contrib<- Contributions_var[, 1:3]

corrplot(dim_contrib, is.corr=FALSE, method="circle",tl.srt=45, tl.col="#003300",
col=brewer.pal(n=9,name="OrRd"),addCoef.col="black")

Contributions_var
```

#Cosinus 2 des vars

```
Cosinus2_Var=round(res.pca$var$cos2,2)

dim_cos2 <- Cosinus2_Var[, 1:3]

corrplot(dim_cos2, is.corr=FALSE, method="circle",tl.srt=45, tl.col="#004400",
col=brewer.pal(n=9,name="PiYG"),addCoef.col="black")

Cosinus2_Var
```

#Cercle de corrélation sur les axes 1 et 2

```
pca_12<-fviz_pca_var(res.pca,axes=c(1,2),col.var="cos2", repel = TRUE)+
  scale_color_gradient2(low="lightsteelblue3",mid="orange",
                        high="red",midpoint=0.65,space="Lab")

pca_12 + ggtitle("Cercle de corrélation Axe 1 et 2")+
theme(plot.title = element_text(hjust = 0.5))
```

#Cercle de corrélation sur les axes 1 et 3

```
pca_13<-fviz_pca_var(res.pca,axes=c(1,3),col.var="cos2", repel = TRUE)+
  scale_color_gradient2(low="lightsteelblue3",mid="orange",
                        high="red",midpoint=0.65,space="Lab")

pca_13 + ggtitle("Cercle de corrélation Axe 1 et 3")+
theme(plot.title = element_text(hjust = 0.5))
```

#Projection de la variable illustrative

#Cercle de corrélation avec la variable illustrative sur les axes 1 et 2

```
res.pca.I=PCA(BaseA[,3:15])
```

```
pca_12_I<-fviz_pca_var(res.pca.I,axes=c(1,2),col.var="cos2", col.quanti.sup = "purple4",
repel = TRUE)+ scale_color_gradient2(low="lightsteelblue3",mid="orange",
high="red",midpoint=0.65,space="Lab")
```

```
pca_12_I + ggtitle("Cercle de corrélation Axe 1 et 2 avec la mortalité")+
theme(plot.title = element_text(hjust = 0.5))
```

#Cercle de corrélation avec la variable illustrative sur les axes 1 et 3

```
pca_13_I<-fviz_pca_var(res.pca.I,axes=c(1,3),col.var="cos2", col.quanti.sup =
"purple4",repel = TRUE)+ scale_color_gradient2(low="lightsteelblue3",mid="orange",
high="red",midpoint=0.65,space="Lab")
```

```
pca_13_I + ggtitle("Cercle de corrélation Axe 1 et 3 avec la mortalité")+
theme(plot.title = element_text(hjust = 0.5))
```

#Etude des individus

#Coordonnées des individus

```
Coordonnees_ind<-round(res.pca$ind$coord,2)
grid.table(Coordonnees_ind)
Coordonnees_ind
```

#Contribution des individus

```
Contribution_ind<-round(res.pca$ind$contrib,2)
grid.table(Contribution_ind)
Contribution_ind
```

#Cos2 des individus

```
Cosinus2_ind <-round(res.pca$ind$cos2,2)
grid.table(Cosinus2_ind)
Cosinus2_ind
```

#Autre méthode pour copier-coller les caractéristiques des individus dans Excel

```
library(Factoshiny)
PCAshiny(base1) #sans la variable illustrative
```

#Projection des individus les plus contributifs

#Axe 1 et 2

```
plot(res.pca, title= "Les 20 pays les plus contributifs", select="contrib 20", col.ind="red")
```

#Axe 1 et 3

```
plot(res.pca,axes=c(1,3), title= "Les 20 pays les plus contributifs", select="contrib  
20",col.ind="red" )
```

#Régression du modèle axe 1 et 2

```
modele<- lm(base2$Mortalité~res.pca$ind$coord[,1]+res.pca$ind$coord[,2])  
summary(modele)
```

#Régression du modèle axe 1 et 3

```
modele1<- lm(base2$Mortalité~res.pca$ind$coord[,1]+res.pca$ind$coord[,3])  
summary(modele1)
```

#Résidus sur le modèle des axes 1 et 2

```
base2$residu=round(residuals(modele),2)  
base2$residu
```

#Résidus sur le modèle des axes 1 et 3

```
base2$residu1=round(residuals(modele1),2)  
base2$residu1
```

#Boxplot et histogramme résidus

```
par(mfrow=c(1,2))  
boxplot(base2$residu, main = "Boxplot résidu axe 1 et 2")  
hist(base2$residu, main="résidu", freq=TRUE, col="gray", cex.main =0.8)
```

```
rosnerTest(base2$residu ,k=4, alpha=0.05)  
rosnerTest(base2$residu1 ,k=4, alpha=0.05)
```

#Skewness et kurtosis sur les résidus du 1^{er} modèle

```
skewness(base2$residu)  
kurtosis(base2$residu)
```

#Skewness et kurtosis sur les résidus du 2^{ème} modèle

```
skewness(base2$residu1)
```

```
kurtosis(base2$residu1)
```

#Test de normalité

```
ks.test(base2$residu,"pnorm",mean(base2$residu),sd(base2$residu))
```

```
ks.test(base2$residu1,"pnorm",mean(base2$residu1),sd(base2$residu1))
```

```
qqnorm(resid(modele))
```

```
qqline(resid(modele), col="red")
```

#homoscédasticité des erreurs

```
bptest(modele)
```

```
bptest(modele1)
```

#non corrélation des erreurs

```
plot(modele)
```

```
plot(modele1)
```

#Pour l'axe 1 et 3

```
par(mfrow=c(1,3))
```

```
boxplot(base2$residu1, main = "Boxplot résidu axe 1 et 3")
```

```
hist(base2$residu1, main="résidu1", freq=TRUE, col="gray", cex.main = 0.8)
```

```
qqnorm(resid(modele1))
```

```
qqline(resid(modele1), col="red")
```

```
ks.test(base2$residu1,"pnorm",mean(base2$residu1),sd(base2$residu1))
```

```
bptest(modele1)
```

#Modèle prédit vs observé axe 1 et 2

```
base2$predict=predict(modele)
```

```
round(base2$predict,2)
```

```
base2$Mortalité
```

```
plot(base2$predict,base2$Mortalité, col=0,main="Mortalité prédit par le modèle et
Mortalité réel", abline(a=0, b=1, col="red"))
text(base2$predict,base2$Mortalité,row.names(base2), cex=.6)
```

#Modèle prédit vs observé axe 1 et 3

```
base2$predict1=predict(modele1)
round(base2$predict1,2)
base2$Mortalité
plot(base2$predict1,base2$Mortalité, col=0,main="Mortalité prédit par le modèle dans
l'axe 1 et 3 et Mortalité réel", abline(a=0, b=1, col="red"))
text(base2$predict1,base2$Mortalité,row.names(base2), cex=.6)
```

#Logit médiane = 25

```
base2$MortalitéN <- ifelse(base2$Mortalité > 25, 1, 0)
base2$MortalitéN=as.numeric(base2$MortalitéN)
base2$MortalitéN
```

#Logit sur les axes 1 et 2

```
lmmodel=glm(base2$MortalitéN~res.pca$ind$coord[,1]+res.pca$ind$coord[,2],family=
binomial)
summary(lmmodel)
```

#logit axe 1 et 3

```
lmmodel1=glm(base2$MortalitéN~res.pca$ind$coord[,1]+res.pca$ind$coord[,3],family
=binomial)
summary(lmmodel1)
```

#Logit Q1 = 12,25

```
summary(base2)
base2$MortalitéN <- ifelse(base2$Mortalité > 12.25, 1, 0)
base2$MortalitéN=as.numeric(base2$MortalitéN)
base2$MortalitéN
```

```
lmmodel=glm(base2$MortalitéN~res.pca$ind$coord[,1]+res.pca$ind$coord[,2],family=
binomial)
```

```
summary(lmmodel)
```

#logit axe 1 et 3

```
lmmodel1=glm(base2$MortalitéN~res.pca$ind$coord[,1]+res.pca$ind$coord[,3],family  
=binomial)
```

```
summary(lmmodel1)
```

#Logit seuil = 20

```
summary(base2)
```

```
base2$MortalitéN <- ifelse(base2$Mortalité > 20, 1, 0)
```

```
base2$MortalitéN=as.numeric(base2$MortalitéN)
```

```
base2$MortalitéN
```

```
lmmodel=glm(base2$MortalitéN~res.pca$ind$coord[,1]+res.pca$ind$coord[,2],family=  
binomial)
```

```
summary(lmmodel)
```

#logit axe 1 et 3

```
lmmodel1=glm(base2$MortalitéN~res.pca$ind$coord[,1]+res.pca$ind$coord[,3],family  
=binomial)
```

```
summary(lmmodel1)
```

VII. Bibliographie

Lien des articles et sites web cités :

Hépatite B chez l'enfant [en ligne], Unicef, disponible sur :

<https://www.unicef.fr/convention-droits-enfants/sante/maladies-infantiles/hepatite-b/> [consulté le 08/11/2023]

Anémie [en ligne] Organisation mondiale de la Santé. Disponible sur :

https://www.who.int/fr/health-topics/anaemia#tab=tab_2 [consulté le 08/11/2023]

L'aide publique au développement (APD) [en ligne], OCDE. Disponible sur :

<https://www.oecd.org/fr/cad/financementpourledeveloppementdurable/normes-financement-developpement/aide-publique-au-developpement.htm> [consulté le 10/11/2023]

Brosseau Anaïs, L'alphabétisation des femmes, un facteur de développement [en ligne], La croix, 08/09/2014. Disponible sur : <https://www.la-croix.com/Actualite/Monde/L-alphabetisation-des-femmes-un-facteur-de-developpement-2014-09-08-1202867> [consulté le 08/11/2023]

Couverture vaccinale [en ligne], OMS, 18/07/2023, disponible sur :

<https://www.who.int/fr/news-room/fact-sheets/detail/immunization-coverage> [consulté le 10/11/2023]

Classement des Etats du monde par taux de mortalité infantile (enfants de moins de 5 ans) [en ligne], Atlasocio, 02/01/2020, Disponible sur :

<https://atlasocio.com/classements/demographie/mortalite/classement-etats-par-taux-de-mortalite-infantile-moins-de-5-ans-monde.php> [consulté le 08/11/2023]

Eau, assainissement et hygiène (EAH) [en ligne], Uncief. Disponible sur :

<https://www.unicef.org/fr/eau-assainissement-hygiene> [consulté le 08/11/2023]

Enfants améliorer leur survie et leur bien-être [en ligne], Organisation mondiale de la Santé, 9/09/2020. Disponible sur :

<https://www.who.int/fr/news-room/fact-sheets/detail/children-reducing-mortality>[consulté le 08/11/2023]

Espérance de vie à la naissance [en ligne], OCDE. Disponible sur :

<https://data.oecd.org/fr/healthstat/esperance-de-vie-a-la-naissance.htm>[consulté le 08/11/2023]

La moitié de la population de la planète n'a pas accès aux services de santé essentiels [en ligne], OMS, 13/12/2017. Disponible sur :

<https://www.who.int/fr/news/item/13-12-2017-world-bank-and-who-half-the-world-lacks-access-to-essential-health-services-100-million-still-pushed-into-extreme-poverty-because-of-health-expenses>[consulté le 08/11/2023]

L'anémie pendant la grossesse : tout savoir [en ligne]May, 13/06/2023. Disponible sur : <https://www.may.app/blog/grossesse/lanemie-pendant-la-grossesse-tout-savoir/#:~:text=Une%20an%C3%A9mie%20s%C3%A9v%C3%A8re%20n'est,essoufflement%20et%20de%20fatigue%20constante> [consulté le 08/11/2023]

Mortalité maternelle [en ligne] Organisation mondiale de la santé, 22/02/2023.

Disponible sur : <https://www.who.int/fr/news-room/fact-sheets/detail/maternal-mortality>[consulté le 08/11/2023]

Ne polluez pas mon avenir ! L'impact environnemental sur la santé infantile [en ligne], Organisation mondiale de la Santé, 21/03/2017. Disponible sur :

<https://www.who.int/fr/publications-detail/WHO-FWC-IHE-17.01> [consulté le 08/11/2023]

ODD 3 : Permettre à tous de vivre en bonne santé et promouvoir le bien-être de tous à tout âge [en ligne], ONU FEMMES. Disponible sur :

<https://www.unwomen.org/fr/news/in-focus/women-and-the-sdgs/sdg-3-good-health-well->

[being#:~:text=D'ici%20%C3%A0%202030%2C%20%C3%A9liminer,000%20naissances%20vivantes%20au%20plus](#) [consulté le 08/11/2023]

Projet afribiota : malnutrition infantile [en ligne], Insitut Pasteur. Disponible sur : <https://www.pasteur.fr/fr/institut-pasteur/institut-pasteur-monde/programmes-internationaux/projet-afribiota-malnutrition-infantile#:~:text=Lutter%20contre%20la%20malnutrition%20infantile,-Un%20enfant%20sur&text=Celle%2Dci%20est%20en%20cause,du%20d%C3%A9veloppement%20physique%20et%20mental> [consulté le 08/11/2023]

Taux de mortalité des enfants de moins cinq ans (pour 1000 naissances vivantes) [en ligne] Organisation mondiale de la santé. Disponible sur : <https://data.who.int/fr/indicators/i/2322814> [consulté le 08/11/2023]

Liens des bases de données:

Dépenses intérieures de santé des administrations publiques (ESEG-D) en pourcentage des dépenses des administrations publiques (ESB) (%) [en ligne],OMS. Disponible sur : <https://data.who.int/fr/indicators/i/B9C6C79> [consulté le 08/11/2023] => Données pour la mortalité, les médecins, les infirmiers, l'anémie, l'hépatite B, le DTC3, Recherche et santé, Aide publique et la maigreur extrême

Espérance de vie à la naissance, total (années) [en ligne] La banque mondiale. Disponible sur : <https://donnees.banquemondiale.org/indicateur/SP.DYN.LE00.IN?most recent year de sc=true> [consulté le 08/11/2023]

Inscriptions à l'école, primaire, filles (%brut) [en ligne] La banque mondiale. Disponible sur : <https://donnees.banquemondiale.org/indicateur/SE.PRM.ENRR.FE> [consulté le 08/11/2023]

People using at least basic drinking water services (% of population) [en ligne] La banque mondiale. Disponible sur :

<https://donnees.banquemondiale.org/indicateur/SH.H2O.BASW.ZS> [consulté le 08/11/2023]

PIB par habitant (\$ US courant) [en ligne] La banque mondiale. Disponible sur :
<https://donnees.banquemondiale.org/indicateur/NY.GDP.PCAP.CD> [consulté le 08/11/2023]

Détails sur les indicateurs :

PIB et ses composantes [en ligne], INSEE, disponible sur :
<https://www.insee.fr/fr/statistiques/2415846> [consulté le 10/11/2023]

PIB/habitant [en ligne], Statista, disponible sur :
<https://fr.statista.com/statistiques/661077/classement-pays-plus-eleve-monde-pib/#:~:text=Le%20PIB%20par%20habitant%20est,%C3%A9lev%C3%A9%20au%20monde%20en%202023%20> [consulté le 10/11/2023]

VIII. Table des matières

Sommaire.....	2
I. Introduction.....	3
II. Analyse détaillée du sujet.....	5
A. Pertinence de la variable à expliquer	5
B. Les variables explicatives et leur corrélation avec la variable illustrative.....	6
1. Les variables explicatives	6
2. Rôle de variables actives et corrélation avec la variable illustrative.....	8
III. Analyse en Composantes Principales ACP	21
A. Analyse statistique des variables	21
B. Valeurs propres et variables : contribution, corrélations, cosinus carrés	26
1. Valeurs propres et nombre d'axes :	26
2. Variables : coordonnées, contribution, cosinus carrés.....	26
C. Cercle de corrélation et variables latentes	28
1. Les cercles de corrélations	28
2. Définition des variables latentes.....	30
3. Projection de la variable illustrative	32
D. Projection des individus sur le plan factoriel, interprétation	34
IV. Corrélation entre la variable à expliquer et les variables latentes	40
A. Régression linéaire multiples	40
B. Relation entre la variable illustrative observée et prédite par la RLM	44
C. Régression logistique binaire	44
V. Conclusion et discussion des résultats	47
VI. Annexes.....	48
A. Tableaux et graphiques supplémentaires	48
B. Base de données	51
C. Script R	52
VII. Bibliographie	63