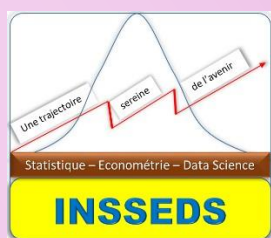


**MINISTRE DE
L'ENSEIGNEMENT
SUPERIEUR ET DE
RECHERCHE SCIENTIFIQUE**



Institut Supérieur de Statistique
D'Econométrie et de Data
Science

REPUBLIQUE DE



CYCLE INGENIEUR
INGENIERIE STATISTIQUE

MINI-PROJET

STATISTIQUE DES VARIABLES QUANTITATIVES ET QUALITATIVES

ETUDE STATISTIQUE POUR UNE ALLOCATION STRATEGIQUE DE RESSOURCES HUMANITAIRES

ANNEE ACADEMIQUE
2024 – 2025

Etudiant

SANA

BOUKARY

**Enseignant –
Encadreur**

**AKPOSSO DIDIER
MARTIAL**

THEME

l'analyse des indicateurs de développement humain et économique des pays. Ce jeu de données inclut des variables essentielles telles que le PIB par habitant, l'espérance de vie, les taux de mortalité infantile, les dépenses de santé, les exportations, les importations, le taux de fertilité, le revenu, et le taux de croissance, entre autres. Ces indicateurs permettent d'analyser les niveaux de développement, les conditions de vie, et les disparités économiques et sociales entre différents pays.

INTRODUCTION GENERALE

Contexte et justification de l'étude

Les inégalités de développement et les crises humanitaires persistent, les organisations non gouvernementales (ONG) jouent un rôle crucial en apportant une aide essentielle aux pays en difficulté. HELP International, une ONG humanitaire internationale, qui a pour mission de combattre la pauvreté et d'intervenir en cas de catastrophe en fournissant des biens de première nécessité et un soutien médical d'urgence. Grâce à une récente collecte de fonds ayant permis de réunir 6 milliards de Fcfa, HELP International doit déterminer comment allouer ces ressources de manière stratégique pour maximiser son impact. Cette étude vise donc à identifier les pays ayant le plus besoin d'assistance en analysant des facteurs socio-économiques et sanitaires.

Problématique

Pour répondre à la mission de HELP International, il est essentiel d'identifier les pays où les besoins humanitaires sont les plus pressants. Le PDG de l'ONG souhaite ainsi se concentrer sur les pays les plus vulnérables, en prenant en compte divers indicateurs de développement. La question centrale est donc : Quels sont les pays qui nécessitent le plus d'aide, et comment les identifier en s'appuyant sur des indicateurs socio-économiques et sanitaires ?

L'objectif de l'étude

L'objectif de cette étude est donc d'optimiser la prise de décision de HELP International en guidant l'allocation de ses ressources aux pays les plus vulnérables, en s'appuyant sur une analyse statistique rigoureuse et des critères socio-économiques ciblés.

Résultats attendus

Cette étude permettra de :

Catégoriser les pays en fonction de leurs caractéristiques socio-économiques et sanitaires.

Identifier et prioriser les pays ayant un besoin urgent d'aide.

Proposer des recommandations basées sur des analyses statistiques, afin de guider HELP International dans ses choix d'allocation de fonds.

Méthodologie

Dans le cadre de cette étude, nous allons adopter une approche statistique pour analyser les indicateurs socio-économiques et sanitaires des pays en difficulté, en vue de déterminer ceux qui nécessitent le plus d'aide. En utilisant le logiciel R, nous allons mener une série d'analyses univariées et bivariées afin de comprendre la distribution de chaque variable et d'identifier les relations significatives entre elles. L'analyse univariée permettra d'observer la tendance centrale et la dispersion des données pour chaque indicateur, tandis que l'analyse bivariable, incluant les corrélations et visualisations croisées, aide à évaluer les interactions entre différents facteurs, tels que la mortalité infantile et le PIB par habitant. Ces étapes sont essentielles pour établir un profil des pays les plus vulnérables et guider les décisions de priorisation pour l'allocation des fonds de HELP International.

Description du jeu de données : dictionnaire des données

Le jeu de données « help international » comprend les variables suivantes : pays : Nom du pays. Enfant mort : Décès d'enfants de moins de 5 ans pour 1000 naissances vivantes. Exportations : Exportations de biens et services par habitant, en pourcentage du PIB par habitant. dep_sante : Dépenses de santé par habitant, en pourcentage du PIB par habitant. Importations : Importations de biens et services par habitant, en pourcentage du PIB par habitant. Revenu : Revenu net par personne. Taux inflation : Taux annuel d'inflation du PIB total. Esperance vie : Espérance de vie à la naissance. Total fertilité : Nombre d'enfants par femme si les taux de fécondité actuels demeurent constants. PIB_par_hab : PIB par habitant, calculé comme le PIB total divisé par la population totale.

Revue de la Littérature

Contexte des facteurs socio-économiques et sanitaires

Les indicateurs socio-économiques et sanitaires jouent un rôle fondamental dans l'évaluation du développement global d'un pays et de la qualité de vie de ses habitants. Dans le cadre des interventions humanitaires, ces facteurs offrent une perspective holistique sur les conditions de vie, incluant la santé, la stabilité économique, et l'accès aux services essentiels. La littérature scientifique souligne que des indicateurs comme le taux de mortalité infantile, l'espérance de vie, et le revenu par habitant sont des révélateurs clés des inégalités structurelles au sein des pays en difficulté. Par exemple, l'Organisation mondiale de la santé (OMS) et la Banque mondiale considèrent que

l'espérance de vie et le taux de mortalité infantile sont parmi les indicateurs les plus pertinents pour évaluer les systèmes de santé et la précarité des conditions de vie.

De même, la stabilité économique, souvent mesurée par le revenu par habitant, le taux d'inflation, et le PIB par habitant, est cruciale pour comprendre la capacité d'un pays à répondre aux besoins de ses citoyens. Une étude réalisée par la Banque mondiale montre que les pays à faible revenu ont souvent des systèmes de santé sous-financés et des infrastructures limitées, ce qui les rend particulièrement vulnérables aux crises. Ces facteurs économiques influent directement sur les capacités des populations à accéder aux soins de santé et aux services de base, amplifiant ainsi les inégalités face aux besoins en aide humanitaire.

Importance des indicateurs dans la priorisation des pays en besoin d'aide

La priorisation des pays pour l'allocation des ressources humanitaires nécessite une analyse rigoureuse de ces indicateurs socio-économiques et sanitaires. Selon l'Organisation des Nations Unies (ONU), une allocation efficace des ressources repose sur une compréhension approfondie des besoins spécifiques de chaque pays, mesurés par des indicateurs précis. Les organisations humanitaires, telles que HELP International, s'appuient souvent sur ces indicateurs pour identifier les pays les plus vulnérables et décider des interventions prioritaires.

Dans ce contexte, le taux de mortalité infantile et l'espérance de vie sont utilisés pour évaluer les conditions de santé d'un pays, tandis que les indicateurs économiques tels que le revenu par habitant et le PIB par habitant aident à comprendre la capacité des pays à faire face aux besoins de leurs populations. Des recherches montrent également que l'inflation et les dépenses de santé influencent directement le bien-être des populations, car elles conditionnent l'accès aux services et la stabilité des économies locales. En classant les pays selon ces indicateurs, les organisations peuvent non seulement cibler leurs interventions, mais aussi mieux anticiper les zones à haut risque et les domaines nécessitant un soutien particulier, ce qui est essentiel pour maximiser l'impact de l'aide humanitaire.

Cette revue de la littérature met en évidence que les indicateurs socio-économiques et sanitaires constituent des outils incontournables pour la priorisation des pays en besoin d'aide, offrant aux organisations humanitaires une base solide pour une prise de décision stratégique et optimisée.

I-ETUDE DE LA BASE DE DONNEES

Dans cette partie nous allons faire l'importation de la base de données dans un premier temps, afficher un aperçu et faire le résumé statistique de la base de données dans un second temps et enfin faire le prétraitement de la base de données.

1-IMPORTATION DE LA BASE DE DONNEES

	PAYS	ENFANT_MORT	EXPORTATIONS	DEP_SANTE
1	AFGHANISTAN	90.2	10.000	7.58
2	ALBANIA	16.6	28.000	6.55
3	ALGERIA	27.3	38.400	4.17
4	ANGOLA	119.0	62.300	2.85
5	ANTIGUA AND BARBUDA	10.3	45.500	6.03

Voici un aperçu de notre base de données.

2-Inspection de la base de données

```
      pays enfant_mort exportations dep_sante importations revenu
162  Uzbekistan      36.3         31.7      5.81         28.5    4240
163   Vanuatu       29.2         46.6      5.25         52.7    2950
164  Venezuela     17.1         28.5      4.91         17.6   16500
165   Vietnam     23.3         72.0      6.84         80.2    4490
166    Yemen      56.3         30.0      5.18         34.4    4480
167    Zambia     83.1         37.0      5.89         30.9    3280
      taux_croissance life_expec total_fertilite pib_par_hab
162          16.50      68.8         2.34         1380
163           2.62      63.0         3.50         2970
164          45.90      75.4         2.47        13500
165          12.10      73.1         1.95         1310
166          23.60      67.5         4.67         1310
167          14.00      52.0         5.40         1460
```

Cette commande nous permet de voir les éléments en tête

La structure de la base de données

```
'data.frame':  167 obs. of  10 variables: $ pays           : Factor w/
167 levels "Afghanistan",...: 1 2 3 4 5 6 7 8 9 10 ...
$ enfant_mort   : num  90.2 16.6 27.3 119 10.3 14.5 18.1 4.8 4.3 39.2 ...
$ exportations  : num  10 28 38.4 62.3 45.5 18.9 20.8 19.8 51.3 54.3 ...
$ dep_sante     : num  7.58 6.55 4.17 2.85 6.03 8.1 4.4 8.73 11 5.88 ...
$ importations  : num  44.9 48.6 31.4 42.9 58.9 16 45.3 20.9 47.8 20.7 .
```

```

$ revenu      : num  1610 9930 12900 5900 19100 18700 6700 41400 43200
$ taux_croissance: num  9.44 4.49 16.1 22.4 1.44 20.9 7.77 1.16 0.873 13.8
$ life_expec   : num  56.2 76.3 76.5 60.1 76.8 75.8 73.3 82 80.5 69.1 .
$ total_fertilite: num  5.82 1.65 2.89 6.16 2.13 2.37 1.69 1.93 1.44 1.92
$ pib_par_hab  : num  553 4090 4460 3530 12200 10300 3220 51900 46900

```

Cette base de données est une data frame qui comporte 167 observations et 10 variables. Les variables sont essentiellement quantitatives.

3-Resume statistique

```

pays      enfant_mort      exportations      dep_sante
Afghanistan      : 1      Min.      : 2.60      Min.      : 0.109      Min.      : 1.810
Albania          : 1      1st Qu.: 7.85      1st Qu.: 23.800      1st Qu.: 4.950
Algeria          : 1      Median : 19.20     Median : 35.000      Median : 6.330
Angola           : 1      Mean    : 37.80     Mean    : 41.353      Mean    : 7.761
Antigua and Barbuda: 1      3rd Qu.: 60.40     3rd Qu.: 51.400      3rd Qu.: 8.865
Argentina        : 1      Max.    :208.00     Max.    :200.000      Max.    :85.100
(Other)          :161      NA's    :4
importations      revenu      taux_croissance      life_expec
Min.      : 0.0659      Min.      : 49.6      Min.      : -4.21      Min.      : 2.04
1st Qu.: 28.8500      1st Qu.: 3170.0      1st Qu.: 1.81      1st Qu.:64.30
Median : 42.9000      Median : 9940.0      Median : 5.45      Median :73.10
Mean    : 45.6223      Mean    : 17023.2      Mean    : 130.49      Mean    :69.30
3rd Qu.: 58.0500      3rd Qu.: 22800.0      3rd Qu.: 11.15      3rd Qu.:76.80
Max.    :174.0000      Max.    :125000.0      Max.    :11400.00      Max.    :82.80

total_fertilite  pib_par_hab
Min.      : 1.15      Min.      : 1.47
1st Qu.: 1.82      1st Qu.: 1310.00
Median : 2.41      Median : 4660.00
Mean    : 4.39      Mean    : 12901.54
3rd Qu.: 4.16      3rd Qu.: 14050.00
Max.    :74.00      Max.    :105000.00

```

Les résultats du résumé statistique, nous donnent pour chaque variable : La moyenne (mean), La valeur minimale (Min.), Le premier quartile (1st Qu), La médiane (Médian), Le troisième quartile (3rd Qu.), La valeur maximale (Max.).

4-PRETRAITEMENT DES DONNEES

Dans le prétraitement des données, nous allons essayer de voir s'ils y'a des doublons dans notre base de données, voir s'il y a des valeur manquantes.

5- Traitement des doublons

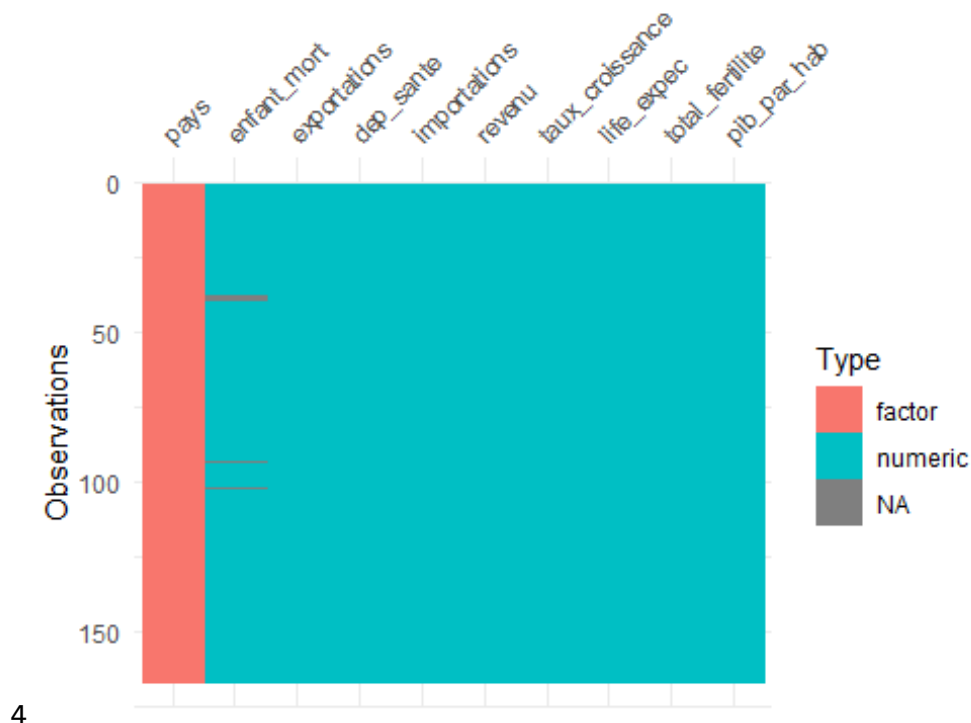
a) vérification des doublons

```
[1] 0
```

Ce résultat nous montre qu'il n'y a aucun doublon dans notre base de données par conséquent la base de données ne nécessite aucun traitement.

b) Traitement des données manquantes

-vérification des données manquantes

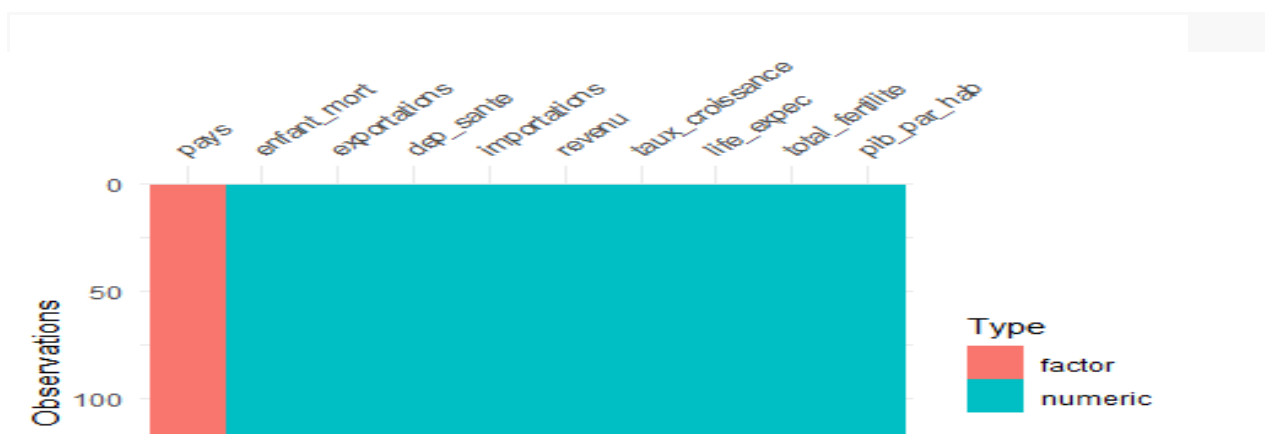


4

[1] 2.39521

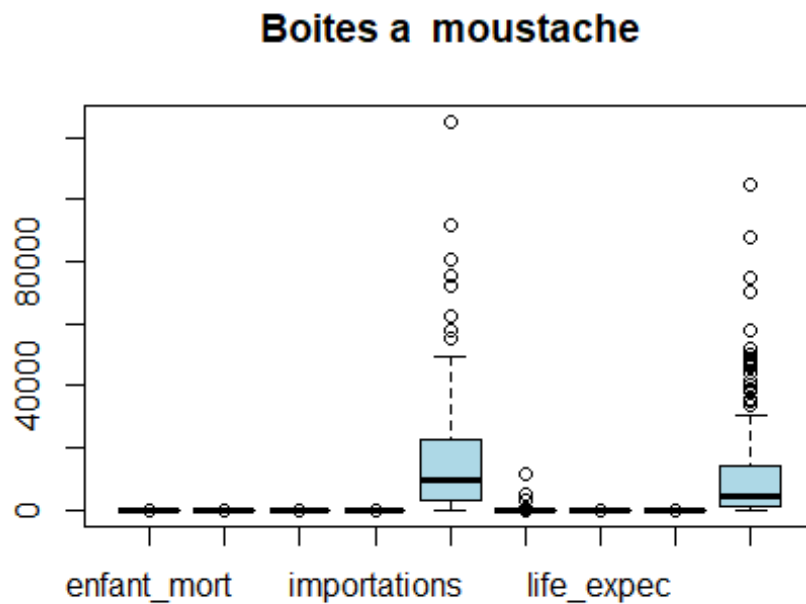
Nous observons 4 valeurs manquantes dans la variable enfant mort, qui correspond a un pourcentage de 2,39. cela nécessite un nettoyage des données et cela pourrait entrainer la suppression des pays potentiellement besoin d'aide. Etant donné que le pays a toute ses modalités et qu'il manque uniquement la variable mort enfant, on va procéder a une imputation par k plus proche voisin (imputation par la médiane) en lieu et place d'une suppression du pays car le PIB par habitant est un facteur déterminant dans notre étude.

d) Traitement des données manquantes

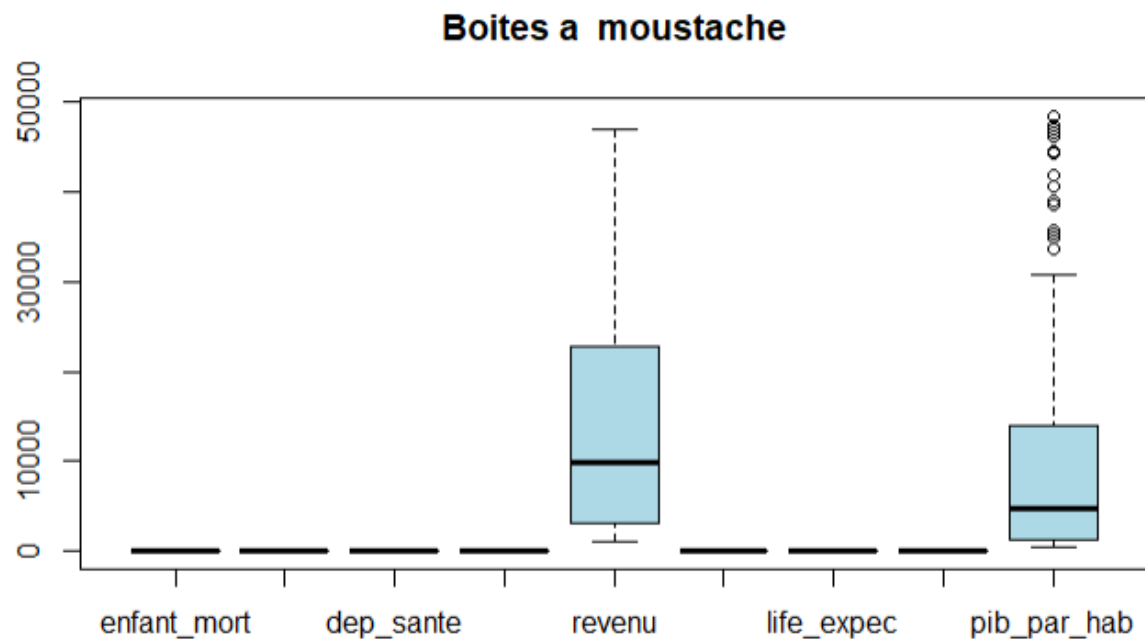


e) Traitement des données extrêmes et aberrantes

Une valeur aberrante est une valeur qui s'écarte fortement des valeurs des autres observations, anormalement faible ou élevée. Dans le cas général elle peut modifier l'interprétation de la moyenne.



Ce graphique nous montre qu'il y a des valeurs extrêmes au niveau de chaque variable. Nous remarquons que les valeurs extrêmes sont plus concentrées sur les variable revenu et pib_par_hab. Passons maintenant aux nettoyages des valeurs extrêmes.



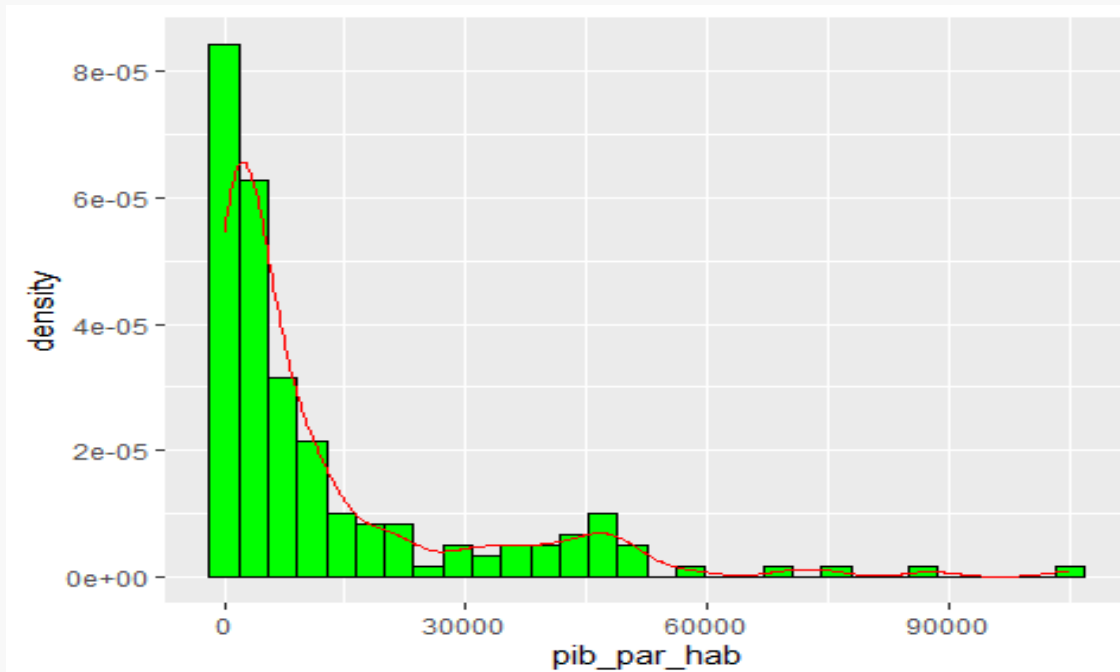
Après traitement des valeurs extrêmes nous passerons maintenant à l'analyse univarié

II-ANALYSE UNIVARIEE

1-ETUDES DES VARIABLE SOCIAUX ECONOMIQUE

1.1-tude de la variable pib_par_hab

	moyenne	minimum	maximum	mod	median	variance	ecart_type
1	12901.54	1.47	105000	2373.82	4660	337351554	18367.13
	coefficient_variation	skewness				interp Kurt	kurtosis
1	142.3639	2.190675	leptokurtique			8.291719	



Histogramme

Interprétation des statistiques pour la variable *PIB par habitant* :

-*Moyenne (12,901.54)* : La moyenne du PIB par habitant est de 12,901.54, ce qui reflète un niveau de richesse moyen par habitant élevé dans l'ensemble des pays, bien que probablement influencé par des valeurs extrêmes (outliers).

-*Minimum (1.47)* : La valeur minimale est de 1.47, montrant qu'il existe des pays où le PIB par habitant est très bas, ce qui témoigne d'une pauvreté extrême.

-*Maximum (105,000)* : La valeur maximale élevée de 105,000 est un outlier significatif, représentant des pays très riches avec un PIB par habitant exceptionnellement haut.

-*Mode (2,373.82)* : Le mode indique la valeur du PIB par habitant la plus fréquente, qui est relativement faible comparée à la moyenne, ce qui suggère que de nombreux pays ont un PIB par habitant bas.

-*Médiane (4,660)* : La médiane est de 4,660, bien en dessous de la moyenne, ce qui indique une asymétrie positive et suggère que la majorité des valeurs du PIB par habitant sont inférieures à la moyenne, influencée par des valeurs élevées.

-*Variance (337,351,554) et Écart-type (18,367.13)* : La variance et l'écart-type élevés indiquent une grande dispersion du PIB par habitant entre les pays, reflétant des inégalités importantes.

-*Coefficient de Variation (18367.13)* : Le coefficient de variation très élevé confirme la grande hétérogénéité du PIB par habitant à travers les pays.

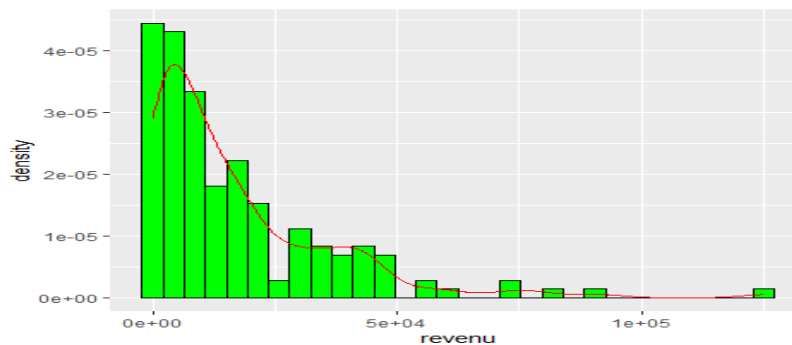
-*Skewness (142.36)* : L'asymétrie est fortement positive, ce qui confirme que la distribution est étirée vers les valeurs élevées, avec quelques pays ayant un PIB par habitant extrêmement haut.

-*Kurtosis (8.29) et Interprétation de la Kurtosis (leptokurtique)* : La kurtosis élevée et la distribution leptokurtique indiquent la présence de valeurs extrêmes (outliers) dans la distribution du PIB par habitant, ce qui renforce l'idée d'inégalités marquées avec quelques valeurs exceptionnellement hautes.

Cette variable est fortement asymétrique avec une majorité de pays ayant un PIB par habitant faible à modéré, tandis que quelques pays montrent un PIB par habitant beaucoup plus élevé.

1.2-etude de la variable revenu

	moyenne	minimum	maximum	mod	median	variance	ecart_type
1	17023.16	49.6	125000	4447.104	9940	374806845	19359.93
	coefficient_variation	skewness	interpkurt	kurtosis			
1		113.727	2.193271	leptokurtique	9.6838		



Histogramme

Interprétation des statistiques descriptives pour la variable *revenu*:

-*Moyenne (17023,16)* : La moyenne indique un revenu par habitant de 17 023,16 \$. Cela montre une forte variation de richesse entre les pays, car cette valeur moyenne est assez élevée par rapport à certains pays.

-*Minimum (49,6) et Maximum (125000)* : Le minimum de 49,6 \$ montre qu'il y a des pays avec des revenus extrêmement faibles par habitant, tandis que le maximum de 125 000 \$ indique des pays avec des revenus très élevés. Cet écart important montre une distribution inégale des revenus dans le monde.

-*Mode (4447,10)* : Le mode, de 4447,10 \$, montre que de nombreux pays ont des revenus par habitant proches de cette valeur. Il est nettement inférieur à la moyenne, ce qui indique que la plupart des pays ont un revenu plus faible, et quelques valeurs élevées influencent la moyenne.

-*Médiane (9940)* : La médiane de 9940 \$ indique que la moitié des pays ont un revenu par habitant inférieur à 9940 \$. Elle est significativement inférieure à la moyenne, ce qui souligne que les revenus plus élevés tirent la moyenne vers le haut.

-*Variance (374 806 845) et Écart type (19359,93)* : La variance très élevée et l'écart type de 19 359,93 \$ indiquent une forte dispersion autour de la moyenne, suggérant des différences significatives dans les revenus des pays.

-*Coefficient de variation (113,73)* : Un coefficient de variation de 113,73 % indique une très grande variabilité relative, ce qui est attendu étant donné les différences de revenu par habitant entre pays riches et pays pauvres.

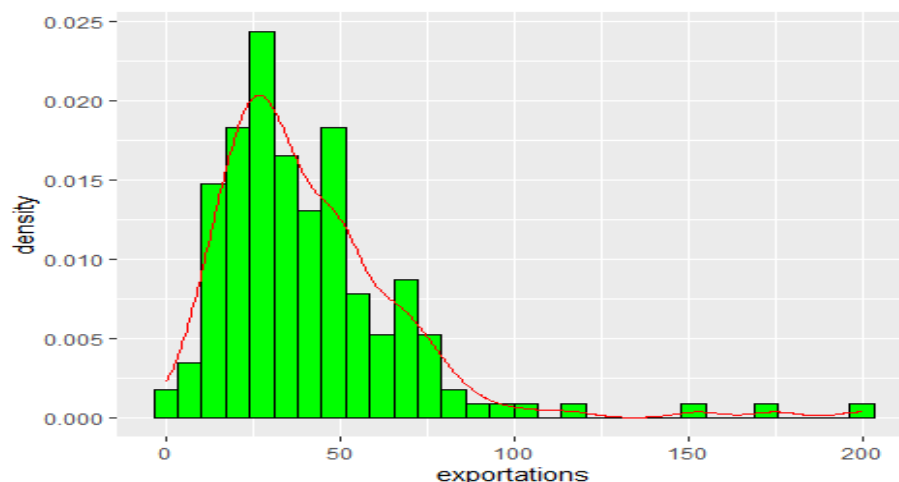
-*Skewness (2,19)* : La skewness de 2,19 montre une asymétrie positive prononcée. Cela signifie que la distribution est fortement biaisée vers des valeurs élevées de revenu, reflétant la présence de pays à revenu très élevé qui influencent la moyenne.

-*Interprétation de la kurtose (leptokurtique) et Kurtosis (9,68)* : La kurtose de 9,68 montre que la distribution est *leptokurtique*. Cela indique une concentration des valeurs près de la moyenne et des extrêmes marqués (valeurs très basses ou très hautes).

En résumé, la variable *revenu* montre une distribution fortement inégale et asymétrique, avec une majorité de pays ayant des revenus par habitant inférieurs à la moyenne, mais avec quelques valeurs extrêmes qui tirent la moyenne vers le haut. Cela met en évidence les écarts de richesse entre pays à faible et à très fort revenu par habitant.

1.3-étude de la variable exportation

```
    moyenne minimum maximum      mod median variance ecart_type
1 41.35329    0.109      200 26.59637      35 780.4267   27.93612
coefficient_variation skewness  interp Kurtosis
1          67.55477 2.353194 leptokurtique 12.07122
```



Interprétation statistiques descriptives pour la variable *exportations* :

-*Moyenne (41,35)* : La moyenne de 41,35 % indique que, en moyenne, les exportations représentent 41,35 % du PIB dans l'échantillon de pays étudiés. C'est une proportion relativement élevée, bien qu'elle puisse masquer une large gamme de valeurs.

-*Minimum (0,109) et Maximum (200)* : Le minimum de 0,109 % montre qu'il y a des pays où les exportations représentent une très faible part du PIB, tandis que le maximum de 200 % suggère des économies extrêmement orientées vers les exportations, où les exportations représentent jusqu'à deux fois le PIB.

-*Mode (26,6)* : Le mode de 26,6 % est inférieur à la moyenne, ce qui pourrait signifier qu'une majorité des pays se trouvent autour d'un niveau d'exportations inférieur à la moyenne de 41,35 %.

-*Médiane (35)* : La médiane de 35 % indique que la moitié des pays ont des exportations représentant au maximum 35 % de leur PIB. La médiane étant inférieure à la moyenne, cela pourrait être le signe d'une asymétrie positive avec quelques valeurs élevées qui augmentent la moyenne.

-*Variance (780,43) et Écart type (27,94)* : La variance de 780,43 et l'écart type de 27,94 révèlent une dispersion significative des pourcentages d'exportations par rapport au PIB, indiquant une grande diversité entre les pays.

-*Coefficient de variation (67,55)* : Le coefficient de variation de 67,55 % montre une variabilité modérée par rapport à la moyenne. Cela indique une distribution avec des valeurs dispersées autour de la moyenne, mais avec une concentration relative.

-*Skewness (2,35)* : La skewness de 2,35 indique une asymétrie positive, signifiant que la distribution des exportations par rapport au PIB est biaisée vers des valeurs

élevées. Cela peut être dû à quelques pays très exportateurs qui influencent la distribution générale.

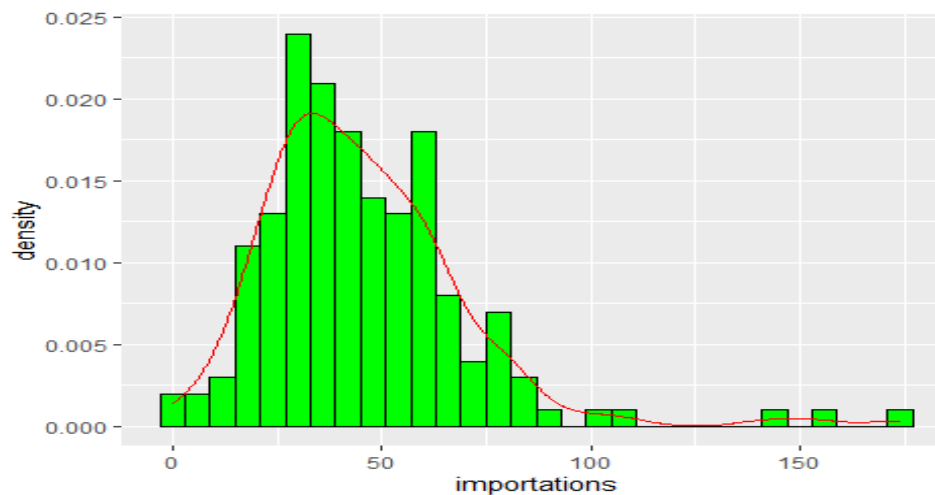
-Interprétation de la kurtose (leptokurtique) et Kurtosis (12,07) : La distribution est leptokurtique, avec une kurtose de 12,07, ce qui indique une distribution présentant des queues épaisses. Cela signifie une forte concentration autour de la moyenne, mais également la présence de valeurs extrêmes.

En résumé, la variable *exportations* montre une répartition assez large, avec quelques valeurs exceptionnellement élevées, ce qui entraîne une asymétrie positive. Cela reflète la diversité des économies, certaines étant très dépendantes des exportations, tandis que d'autres en dépendent très peu. Les valeurs extrêmes dans les exportations influencent la distribution globale.

La majorité des pays ont des exportations relativement faibles (en pourcentage du PIB), mais certains pays montrent des valeurs plus élevées, signalant une forte dépendance économique à l'exportation.

1.4-étude de la variable importation

	moyenne	minimum	maximum	mod	median	variance	ecart_type
1	45.62231	0.0659	174	33.50848	42.9	613.3347	24.7656
	coefficient_variation	skewness	interpkurt	kurtosis			
1		54.28396	1.804001	leptokurtique	9.184421		



Histogramme

Interprétation statistiques descriptives pour la variable *importations*:

-*Moyenne (45,62)* : La moyenne de 45,62 % indique que, en moyenne, les importations représentent 45,62 % du PIB des pays analysés. Cette valeur élevée révèle une forte tendance générale vers l'importation dans les pays inclus.

-*Minimum (0,0659) et Maximum (174)* : Le minimum de 0,0659 % montre qu'il existe des pays où les importations sont presque inexistantes par rapport au PIB, tandis que le maximum de 174 % montre des économies où les importations dépassent le PIB, ce qui peut indiquer une forte dépendance aux importations.

-*Mode (33,51)* : Le mode de 33,51 % est inférieur à la moyenne, ce qui peut signifier qu'un nombre important de pays ont un niveau d'importations autour de cette valeur, tandis que d'autres valeurs plus élevées augmentent la moyenne globale.

-*Médiane (42,9)* : La médiane de 42,9 % indique que la moitié des pays ont des importations représentant 42,9 % ou moins de leur PIB. La proximité de la médiane et de la moyenne indique une répartition assez homogène mais légèrement influencée par des valeurs plus élevées.

-*Variance (613,33) et Écart type (24,77)* : La variance de 613,33 et l'écart type de 24,77 révèlent une dispersion notable des niveaux d'importations dans les pays étudiés, avec des valeurs qui s'écartent significativement de la moyenne.

-*Coefficient de variation (24,77)* : Le coefficient de variation de 24,77 % indique une variabilité modérée par rapport à la moyenne. Cela signifie que, bien que les valeurs diffèrent, elles sont relativement concentrées autour de la moyenne, suggérant une certaine homogénéité.

-*Skewness (1,80)* : La skewness de 1,80 montre une asymétrie positive, ce qui signifie que la distribution est biaisée vers des valeurs plus élevées. Certains pays affichent des importations très élevées, ce qui augmente la moyenne.

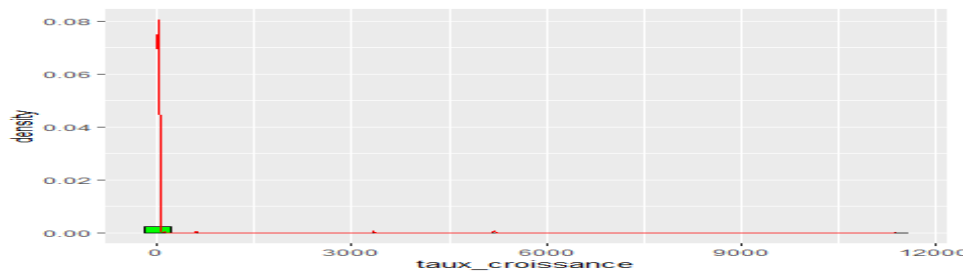
-*Interprétation de la kurtose (leptokurtique) et Kurtosis (9,18)* : La kurtose est de 9,18, qualifiant cette distribution de *leptokurtique*. Cela signifie que la distribution a une concentration accrue autour de la moyenne, mais aussi une tendance aux valeurs extrêmes (avec des « queues » épaisses).

En résumé, la variable *importations* montre une répartition des valeurs relativement homogène, mais influencée par des valeurs extrêmes et une légère asymétrie positive. Certains pays importent beaucoup plus que la moyenne, et ces valeurs élevées influencent la répartition, tandis que la majorité se situe autour de la moyenne de 45,62 %.

Les importations semblent réparties de manière plus uniforme, bien que certains pays montrent une dépendance accrue à l'importation, ce qui peut affecter leur économie de manière significative.

- 1.6-étude de la variable taux de croissance

	moyenne	minimum	maximum	mod	median	variance	ecart_type
1	130.4864	-4.21	11400	11.3721	5.45	997632	998.8153
	coefficient_variation	skewness	interpkurt	kurtosis			
1	765.4556	9.663358	leptokurtique	102.8524			



Interprétation des résultats pour le taux de croissance économique :

-*Moyenne (130.49)* : La moyenne du taux de croissance est de 130, ce qui indique que les valeurs sont généralement élevées, mais elle est probablement influencée par des valeurs extrêmes (outliers).

-*Minimum (-4.21)* : Le minimum indique qu'il existe des pays où le taux de croissance est négatif, suggérant une contraction économique.

-*Maximum (11,400)* : La valeur maximale extrêmement élevée de 11,400 est un outlier significatif, représentant une croissance économique très élevée dans un ou plusieurs pays.

-*Mode (11.37)* : Le mode montre que la valeur la plus fréquente du taux de croissance est de 11.37, suggérant que la plupart des pays connaissent une croissance modérée.

-*Médiane (5.45)* : La médiane, à 5.45, est bien en dessous de la moyenne, ce qui confirme que la distribution est asymétrique avec des valeurs extrêmes vers le haut.

-*Variance (997,632)* et *Écart-type (998.82)* : La variance et l'écart-type élevés confirment une grande dispersion des taux de croissance entre les pays.

-*Coefficient de Variation (998.82)* : Le coefficient de variation élevé confirme l'hétérogénéité significative du taux de croissance entre les pays.

-*Skewness (765.46)* : L'asymétrie positive très marquée (skewness) indique que la distribution est étirée à droite, avec une prédominance de valeurs inférieures et quelques valeurs extrêmement élevées.

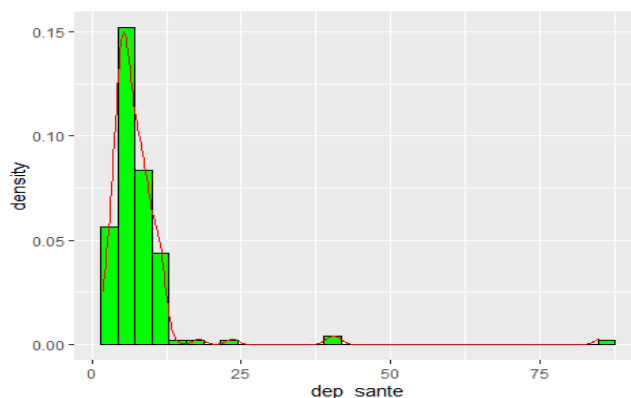
-*Kurtosis (102.85)* et *Interprétation de la Kurtosis (leptokurtique)* : La kurtosis élevée et l'indication "leptokurtique" révèlent une distribution avec de nombreux outliers extrêmes, renforçant l'idée de valeurs rares mais extrêmement élevées.

Le taux de croissance varie largement, avec une majorité autour de valeurs faibles. Certains pays ont des taux de croissance élevés, peut-être liés à des économies émergentes ou à des périodes de reprise.

2-INDICATEURS SANITAIRE

2.1 etude de la variable dep_sante

	moyenne	minimum	maximum	mod	median	variance	ecart_type
1	7.760838	1.81	85.1	5.41692	6.33	58.42598	7.643689
	coefficient_variation	skewness	interpkurt	kurtosis			
1	98.4905	7.23871	leptokurtique	67.74174			



Histogramme

Interprétation statistique descriptive de la variable *dep_sante*:

-*Moyenne (7,76)* : La moyenne de 7,76 % du PIB indique qu'en moyenne, les pays étudiés consacrent cette part de leur PIB aux dépenses de santé. Ce chiffre moyen donne une indication générale de l'engagement financier des pays dans le secteur de la santé.

-*Minimum (1,81) et Maximum (85,1)* : Les valeurs extrêmes montrent une large variation. Le minimum de 1,81 % représente un faible investissement en santé dans certains pays, tandis que le maximum de 85,1 % du PIB est très élevé, indiquant que certains pays consacrent une grande part de leur PIB à la santé, ce qui peut être lié à des besoins sanitaires spécifiques ou des politiques de santé étendues.

-*Mode (5,42)* : Le mode, qui est ici 5,42 %, indique que le pourcentage le plus fréquent des dépenses de santé dans l'échantillon est inférieur à la moyenne, suggérant que la majorité des pays ont des dépenses de santé plus faibles que la moyenne.

-*Médiane (6,33)* : La médiane de 6,33 % signifie que 50 % des pays consacrent 6,33 % ou moins de leur PIB aux dépenses de santé, un chiffre également inférieur à la moyenne, ce qui montre une distribution légèrement biaisée.

-*Variance (58,43) et Écart type (7,64)* : Une variance de 58,43 et un écart type de 7,64 montrent que la dispersion autour de la moyenne est élevée, ce qui reflète la grande diversité dans les budgets de santé entre les pays.

-*Coefficient de variation (98,49)* : Le coefficient de variation proche de 100 % montre que la variabilité des dépenses de santé est élevée par rapport à la moyenne, ce qui est cohérent avec l'ampleur de la gamme de valeurs observées.

-*Skewness (7,24)* : Une skewness de 7,24 indique une forte asymétrie positive, ce qui signifie que la distribution est fortement biaisée vers des valeurs très élevées. Cette asymétrie positive suggère qu'il existe un petit nombre de pays ayant des dépenses de santé exceptionnellement élevées, ce qui tire la distribution vers la droite.

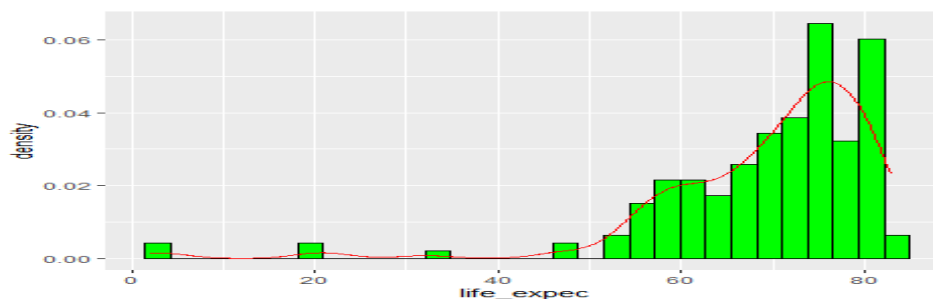
-*Interprétation de la kurtose (leptokurtique) et Kurtosis (67,74)* : La distribution est *leptokurtique*, avec une kurtose élevée de 67,74, ce qui signifie qu'elle présente des queues beaucoup plus épaisses que la normale. Cela confirme la présence de valeurs extrêmes (outliers), indiquant que quelques pays se démarquent par des pourcentages de dépenses de santé extrêmement élevés.

En résumé, les dépenses de santé montrent une forte variabilité entre les pays, avec une distribution très asymétrique et leptokurtique. Quelques pays dépensent des parts très élevées de leur PIB en santé, créant des extrêmes qui influencent les mesures de dispersion et d'asymétrie dans l'ensemble de l'échantillon.

Les dépenses de santé en pourcentage du PIB sont concentrées dans des valeurs basses à moyennes, ce qui pourrait indiquer des limitations dans les investissements en santé dans de nombreux pays.

2.2-etude de la variable life_expec

moyenne	minimum	maximum	mod	median	variance	ecart_type
1	69.29844	2.04	82.8	75.98564	73.1	160.6303
coefficient_variation	skewness	interpkurt	kurtosis			
1	18.28901	-2.593863	leptokurtique	12.61935		



Histogramme

Interprétation des statistiques pour la variable *espérance de vie (life_expec)* :

-*Moyenne (69.3 ans)* : L'espérance de vie moyenne est de 69,3 ans, ce qui indique un niveau de vie et de santé globalement favorable dans les pays étudiés.

-*Minimum (2.04 ans)* : La valeur minimale de 2.04 ans est un outlier significatif, probablement dû à des erreurs de saisie ou à des situations exceptionnelles. Une valeur aussi basse est irréaliste et devrait être examinée pour valider sa précision.

-*Maximum (82.8 ans)* : La valeur maximale est de 82,8 ans, ce qui est en ligne avec l'espérance de vie dans les pays les plus développés.

-*Mode (75.98 ans)* : Le mode, 75.98 ans, est proche de la médiane et au-dessus de la moyenne, ce qui montre qu'une grande proportion de pays ont une espérance de vie relativement élevée.

-*Médiane (73.1 ans)* : La médiane de 73,1 ans est légèrement plus élevée que la moyenne, ce qui signifie qu'environ la moitié des pays ont une espérance de vie supérieure à 69,3 ans.

-*Variance (160.63) et Écart-type (12.67)* : La variance et l'écart-type sont modérés, montrant une dispersion modérée autour de la moyenne et indiquant des variations d'espérance de vie entre les pays.

-*Coefficient de variation (12.67)* : Le coefficient de variation, modéré, confirme une variation raisonnable autour de la moyenne, indiquant une certaine homogénéité dans l'espérance de vie entre pays, malgré quelques écarts notables.

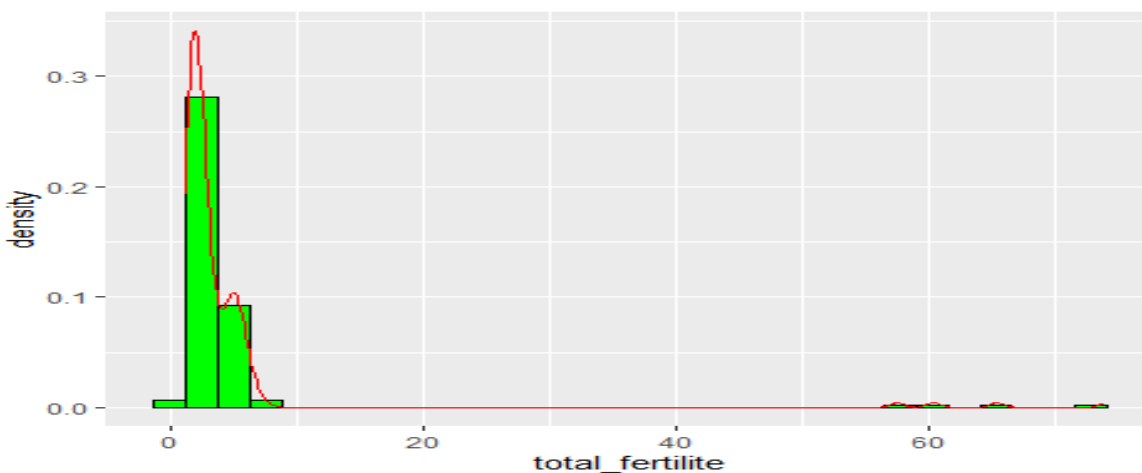
-*Skewness (-2.59)* : L'asymétrie est fortement négative, ce qui signifie que la majorité des pays ont une espérance de vie élevée, avec quelques valeurs extrêmement basses tirant la moyenne vers le bas.

-*Kurtosis (12.62) et interprétation de la Kurtosis (leptokurtique)* : La kurtosis élevée indique une distribution leptokurtique, avec des valeurs extrêmes (outliers), notamment celles très basses, contribuant à cette forte leptokurticité.

La distribution est assez concentrée vers des valeurs moyennes à élevées. Cela reflète les progrès dans la santé globale, bien que certains pays aient encore une espérance de vie inférieure.

2.3- étude de la variable total fertilité

	moyenne	minimum	maximum	mod	median	variance	ecart_type
1	4.390359	1.15	74	1.992228	2.41	91.80082	9.581274
	coefficient_variation	skewness	interpkurt	kurtosis			
1	218.2344	6.102264	leptokurtique	39.64946			



Histogramme

Interprétation des statistiques descriptives pour la variable *total_fertilité*:

-*Moyenne (4,39)* : La moyenne de 4,39 enfants par femme indique le nombre moyen d'enfants qu'une femme est censée avoir au cours de sa vie dans l'échantillon de pays étudiés. Ce chiffre donne une idée générale des niveaux de fécondité, bien qu'il cache des variations significatives.

-*Minimum (1,15) et Maximum (74)* : Le minimum de 1,15 enfants par femme montre qu'il existe des pays avec des taux de fécondité très bas. À l'inverse, le maximum de

74 est extrêmement élevé et peut être le résultat d'une erreur ou d'une donnée extrême dans l'échantillon.

-*Mode (1,99)* : Le mode, à 1,99 enfants par femme, est inférieur à la moyenne et pourrait indiquer que la plupart des pays ont des taux de fécondité plus faibles, près de 2 enfants par femme, dans cette distribution.

-*Médiane (2,41)* : La médiane est de 2,41 enfants par femme, ce qui signifie que la moitié des pays ont des taux de fécondité inférieurs ou égaux à cette valeur. La médiane étant bien en dessous de la moyenne, cela suggère que la distribution est biaisée positivement avec des valeurs plus élevées qui augmentent la moyenne.

-*Variance (91,80) et Écart type (9,58)* : La variance de 91,80 et l'écart type de 9,58 montrent une dispersion importante des taux de fécondité dans l'échantillon, confirmant une forte variabilité entre les pays.

-*Coefficient de variation (218,23)* : Le coefficient de variation de 218,23 % indique une variabilité très élevée par rapport à la moyenne. Cela confirme que les valeurs du taux de fécondité sont extrêmement dispersées autour de la moyenne.

-*Skewness (6,10)* : La skewness de 6,10 montre une forte asymétrie positive, ce qui signifie que la distribution est biaisée vers des valeurs exceptionnellement élevées. Cette asymétrie indique la présence de quelques pays avec des taux de fécondité beaucoup plus élevés que la majorité des autres.

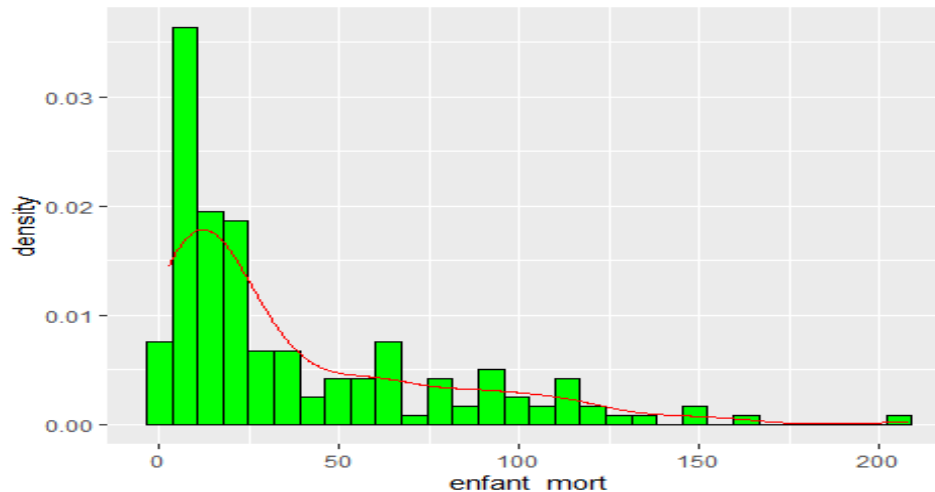
-*Interprétation de la kurtose (leptokurtique) et Kurtosis (39,65)* : La distribution est *leptokurtique*, avec une kurtose de 39,65, ce qui signifie qu'elle présente des queues épaisses, témoignant de la présence de valeurs extrêmes. Cela indique une concentration autour de la moyenne avec quelques valeurs très éloignées.

En résumé, la variable *total_fertilite* présente une grande variabilité, avec une distribution asymétrique positive marquée par quelques valeurs extrêmes. Cela reflète des différences significatives dans les taux de fécondité entre les pays, certains ayant des valeurs très élevées qui influencent les statistiques globales.

La fertilité est élevée dans plusieurs pays, mais tend à être modérée à faible dans la plupart des cas, ce qui peut refléter des tendances de développement démographique.

2.3-étude de la variable enfant mort

	moyenne	minimum	maximum	mod	median	variance	ecart_type
1	37.3509	2.6	208	12.4072	19.2	1589.284	39.86583
	coefficient_variation		skewness		interpkurt		kurtosis
	1		106.7332		1.510813 leptokurtique		4.947671



Histogramme

Interprétation statistique de la variable *enfant_mort* :

-*Moyenne (37,35)* : La moyenne de 37,35 décès d'enfants pour 1 000 naissances indique le niveau moyen de mortalité infantile dans l'ensemble des pays étudiés. Cette valeur suggère que, globalement, la mortalité infantile reste une problématique importante dans certains pays.

-*Minimum (2,6) et Maximum (208)* : Les extrêmes de la distribution montrent des variations importantes dans la mortalité infantile entre les pays. Le minimum de 2,6 représente un niveau très faible, tandis que le maximum de 208 indique que certains pays connaissent des niveaux de mortalité infantile extrêmement élevés.

-*Mode (12,41)* : La valeur modale de 12,41 suggère que le niveau le plus fréquent de mortalité infantile parmi les pays se situe autour de ce chiffre, bien en dessous de la moyenne, indiquant que la majorité des pays ont des niveaux de mortalité infantile inférieurs à la moyenne.

-*Médiane (19,2)* : La médiane de 19,2 indique que 50 % des pays ont un taux de mortalité infantile inférieur ou égal à cette valeur. Cela montre une asymétrie dans la distribution, où les valeurs élevées augmentent la moyenne au-dessus de la médiane.

-*Variance (1589,28) et Écart type (39,87)* : L'écart type élevé (39,87) et la variance importante (1589,28) montrent une dispersion considérable des valeurs autour de la moyenne, soulignant une forte variabilité entre les pays en matière de mortalité infantile.

-*Coefficient de variation (106,73)* : Avec un coefficient de variation supérieur à 100 %, cela indique que la dispersion des valeurs est très grande par rapport à la moyenne, ce qui est typique pour des données ayant une large gamme de valeurs.

-*Skewness (1,51)* : La skewness, ou asymétrie, de 1,51 montre une distribution légèrement asymétrique positive, indiquant que la mortalité infantile est biaisée vers des valeurs élevées. Certains pays ont des taux de mortalité infantile élevés, ce qui tire la distribution vers la droite.

-*Interprétation de la kurtose (leptokurtique) et Kurtosis (4,95)* : La distribution est *leptokurtique*, signifiant qu'elle présente des queues plus épaisses que la normale, suggérant la présence de valeurs extrêmes (outliers) dans la distribution de la mortalité infantile. La kurtose de 4,95 confirme cette leptokurticité, suggérant qu'un petit nombre de pays se démarque avec des niveaux de mortalité infantile exceptionnellement élevés.

En résumé, les données sur la mortalité infantile révèlent une grande variabilité entre les pays, avec une asymétrie positive et une distribution leptokurtique. Ces statistiques montrent que certains pays enregistrent des taux de mortalité infantile bien plus élevés que la moyenne, accentuant la disparité dans la distribution des taux de mortalité infantile.

Cette variable présente une large distribution, avec une concentration plus élevée de valeurs faibles à moyennes. Cela indique que certains pays ont des taux de mortalité infantile particulièrement élevés, mais la majorité a un taux plus modéré.

Les disparités observées, en particulier pour le `pib_par_hab`, l'espérance de vie et la mortalité infantile, montrent que certains pays sont bien plus vulnérables. Les pays avec des taux de mortalité infantile élevés, un PIB faible et une espérance de vie réduite devraient être priorisés pour une allocation optimale des ressources.

Analyse des Indicateurs de Vulnérabilité pour la Priorisation des Pays Nécessiteux : Pour identifier les pays les plus nécessiteux et aider l'aide internationale à prioriser ses interventions, il est pertinent de sélectionner des variables qui reflètent les besoins essentiels et la vulnérabilité socio-économique des pays. En fonction de l'analyse univariée et des caractéristiques des données, voici les variables clés qui semblent les plus pertinentes :

1. *enfant_mort* : Ce taux représente la mortalité infantile, un indicateur crucial pour mesurer la santé et le bien-être des enfants dans un pays. Un taux élevé de mortalité infantile peut signaler des problèmes d'accès aux soins de santé, de nutrition, et d'hygiène. Les pays avec des valeurs élevées dans cette variable devraient être prioritaires.
2. *dep_sante* : La part du PIB consacrée aux dépenses de santé est un indicateur de l'investissement d'un pays dans la santé publique. Les pays avec des niveaux très faibles de dépenses de santé peuvent avoir un système de santé fragile, nécessitant une aide pour faire face aux besoins de leur population.
3. *total_fertilite* : Le taux de fécondité est souvent corrélé avec le développement socio-économique et l'accès à l'éducation et aux services de planification familiale.

Les pays avec un taux de fécondité très élevé pourraient avoir des besoins accrus en éducation et en services de santé maternelle et infantile.

4. *revenu* : Le revenu moyen par habitant est un indicateur direct du niveau de richesse d'un pays. Les pays avec un faible revenu moyen par habitant sont souvent plus vulnérables et en besoin d'assistance économique et sociale.
5. *life_expec* : L'espérance de vie à la naissance est un bon indicateur du niveau de santé général d'une population. Un faible niveau d'espérance de vie peut indiquer un besoin de soutien dans divers domaines, y compris la santé publique, la nutrition et les conditions de vie.
6. *pib_par_hab* : Le PIB par habitant permet de mesurer la richesse économique relative d'un pays. Les pays avec un PIB par habitant bas peuvent bénéficier de programmes d'aide pour soutenir la croissance économique et réduire la pauvreté.
7. *taux_croissance* : Ce taux de croissance démographique peut également être important. Un taux élevé de croissance démographique dans un contexte de pauvreté peut aggraver les difficultés d'un pays à fournir les services de base à sa population croissante.

Priorisation

En combinant ces variables, nous pourrions établir un indice de vulnérabilité pour chaque pays basé sur des critères de santé (*enfant_mort*, *dep_sante*, *life_expec*), de fertilité et de pression démographique (*total_fertilite*, *taux_croissance*), et de situation économique (*revenu*, *pib_par_hab*). Les pays avec les scores de vulnérabilité les plus élevés seraient priorisés pour recevoir une aide en premier.

NB: Dans le cadre de notre étude visant à identifier les pays les plus nécessiteux afin que l'aide internationale puisse leur apporter une aide financière, une *analyse bivariée* pourrait être très pertinente. Elle permettrait de comprendre les relations entre les différentes variables sélectionnées après notre analyse univariée et de voir comment elles interagissent, ce qui pourrait affiner votre sélection des pays en identifiant des combinaisons de facteurs qui indiquent une vulnérabilité accrue.

Variables pour l'analyse bivariée

1. *Enfant_mort* et *revenu* : Le taux de mortalité infantile est souvent plus élevé dans les pays à faibles revenus. La corrélation entre ces deux variables pourrait indiquer si un faible revenu est lié à des conditions de vie précaires pour les enfants.
2. *Enfant_mort* et *dep_sante* : Une relation inverse est souvent attendue entre le taux de mortalité infantile et les dépenses en santé. Une corrélation négative pourrait souligner l'impact des investissements en santé sur la réduction de la mortalité infantile.

3. *Life_expec* et *dep_sante* : Une analyse de la corrélation entre l'espérance de vie et les dépenses en santé pourrait montrer dans quelle mesure les investissements en santé influencent la durée de vie moyenne des populations dans les différents pays.
4. *Taux_croissance* et *pib_par_hab* : La corrélation entre la croissance économique et le PIB par habitant permettrait d'explorer dans quelle mesure le développement économique influence le niveau de vie.
5. *Total_fertilite* et *revenu* : Les pays à plus haut taux de fécondité tendent souvent à avoir des revenus plus bas, ce qui peut aggraver la vulnérabilité économique. Cette corrélation pourrait donc apporter des informations sur les défis économiques de certains pays.

Variables pertinentes pour les corrélations

Parmi ces variables, “*enfant mort*” pourrait être considérée comme la variable pertinente, car elle représente un indicateur direct de la précarité et des conditions de vie d'une population. Le taux de mortalité infantile est souvent influencé par une combinaison de facteurs, notamment le revenu, les dépenses en santé, l'espérance de vie, et le niveau d'éducation ou de fertilité. Ainsi, cette variable pourrait servir de référence centrale pour identifier les corrélations qui révèlent les pays nécessitant une aide prioritaire.

En conclusion, l'analyse bivariée apportera une compréhension plus fine des interactions entre les différents indicateurs de vulnérabilité et pourra ainsi nous guider dans notre processus de sélection des pays à prioriser.

II-ANALYSE BIVARIÉE

1-critères de santé

VÉRIFICATION DE LA LIAISON DE CORRÉLATION ENTRE: “*enfant_mort* & *dep_sante*”

[1] -0.3667752

1.2. Coefficient de Corrélation de Spearman : -0.3667752

- *Interprétation* : Le coefficient de Spearman mesure la corrélation entre les rangs des deux variables, ce qui permet de détecter une relation monotone (qu'elle soit linéaire ou non). Une valeur de -0.367 indique une *corrélation négative faible*. Cela pourrait signifier que, de manière générale, les pays avec des dépenses de santé plus élevées ont un taux de mortalité infantile légèrement plus bas. Cependant, la corrélation étant faible, l'influence des dépenses de santé reste limitée et pourrait être affectée par d'autres facteurs.

Enfant_mort et revenu

[1] -0.8337801

1.3- Coefficient de Corrélation de Spearman : -0.8337801

- *Interprétation* : Le coefficient de Spearman est négatif et assez élevé en valeur absolue (-0.834), ce qui indique une *corrélation négative modérée à forte* entre le revenu et la mortalité infantile. Ce résultat suggère que, de manière générale, les pays avec un revenu plus élevé ont tendance à avoir un taux de mortalité infantile plus bas. Cette relation monotone inverse est plus forte et montre que les rangs des revenus et de la mortalité infantile tendent à être inversement liés.

Conclusion Les résultats montrent une relation inverse modérée entre le revenu et la mortalité infantile, ce qui suggère que les pays avec des revenus plus élevés ont généralement des taux de mortalité infantile plus bas. la corrélation linéaire forte (Pearson) indique que cette relation est purement linéaire. Ces résultats suggèrent que le revenu est un facteur pertinent et déterminant, pour expliquer les différences dans la mortalité infantile. Pour catégoriser les pays nous allons utiliser deux facteurs déterminant qui sont: le facteur mort enfant ; le revenu par habitant.

#CATEGORISATIONS DES PAYS :

pays <fctr>	revenu <dbl>	enfant_mort <dbl>	score_composite <dbl>	categorie <chr>
Liberia	700.0	89.3	653.3023	FAIBLE
Burundi	764.0	93.6	715.0537	FAIBLE
Niger	814.0	123.0	749.6795	FAIBLE
Central African Republic	888.0	149.0	810.0833	FAIBLE
Mozambique	918.0	101.0	865.1840	FAIBLE
Malawi	1030.0	90.5	982.6748	FAIBLE
Guinea	1190.0	109.0	1133.0005	FAIBLE
Sierra Leone	1220.0	160.0	1136.3311	FAIBLE
Togo	1210.0	90.3	1162.7793	FAIBLE
Rwanda	1350.0	63.6	1316.7416	FAIBLE

1-10 of 167 rows

Previous **1** 2 3 4 5 6 ... 17 Next

voici la visualisations des pays par catégories en fonction de leur niveau de précarité.

CONCLUSION GENERALE : Cette étude a permis de classer les pays les plus vulnérables au moins vulnérables afin d'aider le PDG à optimiser la prise de décision de HELP International en guidant l'allocation de ses ressources aux pays les plus vulnérables. Nous avons grâce aux logiciels Rstudio pu mettre en œuvre les techniques de traitement des données et de statistiques descriptives et d'analyse de données vue en cours. Au cours de notre étude nous avons remarqué que le poids pèse plus que d'autres variables.

On pourrait utiliser plus de variables afin d'avoir des résultats plus pertinents.

ANEXES : SOURCE DE CODE R

```
help_international <- read.csv("D:/projet de l'ecole/help_international.csv",  
sep=";", stringsAsFactors=TRUE)  
print(help_international)#extraire la base de donnée
```

```
tail(help_international)  
summary(help_international) #resume statistiques
```

```

sum(duplicated(help_international)) #verification des doublons

library(visdat)
vis_dat(help_international)#affichage des valeur manquante

sum(!complete.cases(help_international))#afficher le nombre de valeur
manquante

## [1] 4

sum(!complete.cases(help_international))/nrow(help_international)*100

help_international$enfant_mort[is.na(help_international$enfant_mort)]=median(
help_international$enfant_mort,na.rm=TRUE)
vis_dat(help_international) #traitement des valeurs manquantes

afficher_boites_a_moustache <- function(dataframe) {
  # Selectionner uniquement les colonnes numeriques
  colonnes_numeriques <- sapply(dataframe, is.numeric)

  if (sum(colonnes_numeriques) > 0) {
    # CrC)er un graphique de boites a moustache pour chaque colonne
    numC)rique
    boxplot(dataframe[, colonnes_numeriques], main="Boites a moustache",
col="lightblue", border="black")
  } else {
    cat("Aucune colonne numerique C afficher.\n")
  }
}

afficher_boites_a_moustache(help_international)

library(dplyr)

library(moments)
library(RVAideMemoire)

## *** Package RVAideMemoire v 0.9-83-7 ***

sana_resume_vecteur <- function(data, vecteur) {
  data %>%
    summarise(
      moyenne = mean(.data[[vecteur]]),
      minimum = min(.data[[vecteur]]),
      maximum = max(.data[[vecteur]]),
      mod = mod(.data[[vecteur]]), median = median(.data[[vecteur]]),
      variance = var(.data[[vecteur]]),
      ecart_type = sd(.data[[vecteur]]),
      coefficient_variation = cv(.data[[vecteur]]),
      skewness = skewness(.data[[vecteur]]),
      interpkurt = if_else(skewness(.data[[vecteur]]) < 0,

```

```

        'distribution étalée à gauche',
        'distribution étalée à droite'),
    kurtosis = kurtosis(.data[[vecteur]]),
    interpkurt = if_else(kurtosis(.data[[vecteur]]) < 3,
        'platikurtique',
        'leptokurtique'))
}

library(ggplot2)
sana_resume_vecteur(help_international, "pib_par_hab")

ggplot(help_international, aes(x = pib_par_hab)) +
  geom_histogram(aes(y = ..density..), bins = 30, fill =
'green', color="black") +
  geom_density( alpha=.2, color = 'red')

ggplot(help_international, aes(x = revenu)) +
  geom_histogram(aes(y = ..density..), bins = 30, fill =
'green', color="black") +
  geom_density( alpha=.2, color = 'red')

sana_resume_vecteur(help_international, "revenu")
sana_resume_vecteur(help_international, "exportations")
sana_resume_vecteur(help_international, "importations")

ggplot(help_international, aes(x = exportations)) +
  geom_histogram(aes(y = ..density..), bins = 30, fill =
'green', color="black") +
  geom_density( alpha=.2, color = 'red')

ggplot(help_international, aes(x = importations)) +
  geom_histogram(aes(y = ..density..), bins = 30, fill =
'green', color="black") +
  geom_density( alpha=.2, color = 'red')

ggplot(help_international, aes(x = taux_croissance)) +
  geom_histogram(aes(y = ..density..), bins = 30, fill =
'green', color="black") +
  geom_density( alpha=.2, color = 'red')

ggplot(help_international, aes(x = life_expec)) +
  geom_histogram(aes(y = ..density..), bins = 30, fill =
'green', color="black") +
  geom_density( alpha=.2, color = 'red')

sana_resume_vecteur(help_international, "enfant_mort")

```

```

sana_resume_vecteur(help_international,"taux_croissance")

correlation_spearman <-
cor(help_international$enfant_mort,help_international$dep_sante, method =
"spearman")
print(correlation_spearman)

correlation_spearman <-
cor(help_international$enfant_mort,help_international$revenu, method =
"spearman")
print(correlation_spearman)

# Charger le package dplyr
library(dplyr)

# Calculer la corrélation entre le revenu et la mortalité infantile
correlation <- cor(help_international$revenu, help_international$enfant_mort,
method = "pearson", use = "complete.obs")

# Vérifier la corrélation
cat("Corrélation entre revenu et mortalité infantile :", correlation, "\n")

## Corrélation entre revenu et mortalité infantile : -0.5067766

# Calculer le score composite pour chaque pays en tenant compte de la
corrélation
help_international <- help_international %>%
  mutate(score_composite = revenu - (enfant_mort * abs(correlation)))

# Calculer les quantiles une seule fois pour stabiliser les seuils
quantile_33 <- quantile(help_international$score_composite, probs = 0.33,
na.rm = TRUE)
quantile_66 <- quantile(help_international$score_composite, probs = 0.66,
na.rm = TRUE)

# Créer le classement des pays par catégories en utilisant des quantiles
fixes
classement_pays <- help_international %>%
  mutate(
    # Catégoriser les pays en fonction des quantiles du score composite
    categorie = case_when(
      score_composite < quantile_33 ~ "FAIBLE",
      score_composite >= quantile_33 & score_composite < quantile_66 ~
"MOYEN",
      TRUE ~ "FORT"
    )
  ) %>%
  arrange(score_composite) %>% # Classer les pays en ordre croissant de
score_composite

```

```
select(pays, revenu, enfant_mort, score_composite, categorie) #
```

Sélectionner les colonnes pertinentes