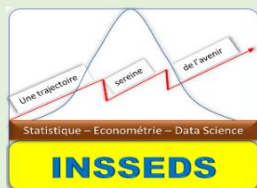


MINISTRE DE L'ENSEIGNEMENT  
SUPERIEUR ET DE RECHERCHE  
SCIENTIFIQUE



Institut Supérieur de  
Statistique  
d'Econométrie et de Data

REPUBLIQUE DE COTE  
D'IVOIRE

Union – Discipline - Travail



**MASTER 2**

**STATISTIQUE – ECONOMETRIE – DATA SCIENCE**

**Mini-projet**  
**Statistique Descriptive**  
**EVOLUTION DU PRIX DU PETROLE AU NIGERIA DE**  
**1960 A 2019**

**2022 – 2023**

**Etudiant**  
**SAME Adama**

**Enseignant – Encadreur**  
**AKPOSSO Didier Martial**

## AVANT-PROPOS

Cher lecteur,

C'est avec un grand enthousiasme que je vous présente ce mini-projet, fruit de plusieurs semaines de réflexion, de recherche et de travail acharné. Ce mini-projet est le résultat de mon engagement envers l'apprentissage, la créativité et la mise en pratique des connaissances acquises.

L'objectif de ce mini-projet est de démontrer ma capacité à appliquer les compétences et les concepts que j'ai acquis dans le cadre de l'obtention du diplôme de **Master Professionnel en Statistique – Econométrie – Machine Learning à l'Institut Supérieur de Statistique d'Econométrie et de Data Science (INSEDS)**. Le sujet abordé dans le cadre de ce mini-projet me semble à la fois stimulant et pertinent.

Au cours de ce travail, je vais aborder de manière descriptive **l'évolution du prix du pétrole au Nigéria de 1960 à 2019** en appliquant les **méthodes de la statistique descriptive aux variables d'intérêt**. Je vais également présenter des recommandations et des conclusions basées sur les données et les informations recueillies.

Je tiens à remercier M. AKPOSSO Didier Martial, Directeur des études et Encadreur de ladite formation, les enseignants de l'INSEDS, ma famille et les amis qui m'ont soutenu tout au long de la rédaction de ce mini-projet en m'apportant leur expertise, leurs conseils et leur encouragement.

J'espère que ce mini-projet sera une source d'inspiration pour ceux qui souhaitent explorer davantage le domaine pétrolier au Nigéria et contribuer à l'enrichissement des connaissances dans ce domaine spécifique.

Enfin, je vous invite à parcourir ce document avec attention, en espérant qu'il vous apportera une vision claire et détaillée de mon travail, tout en suscitant votre intérêt et votre réflexion.

Merci de prendre le temps de découvrir ce mini-projet. Je vous souhaite une agréable lecture.

Cordialement

## TABLE DES MATIERES

<b>AVANT-PROPOS .....</b>	<b>2</b>
<b>INTRODUCTION GENERALE .....</b>	<b>5</b>
<b>PREMIERE PARTIE : PRETRAITEMENT DES DONNEES.....</b>	<b>5</b>
<b>I- DESCRIPTION DU JEU DE DONNEES : dictionnaire des données.....</b>	<b>5</b>
<b>II- TRAITEMENT DES VALEURS MANQUANTES.....</b>	<b>7</b>
<b>III- TRAITEMENT DES VALEURS EXTREMES ET ABERRANTES .....</b>	<b>9</b>
<b>DEUXIEME PARTIE : ANALYSE UNIVARIEE.....</b>	<b>10</b>
<b>I- ANALYSE DES VARIABLES QUANTITATIVES .....</b>	<b>10</b>
1- Rappel du fondement théorique des indicateurs statistiques .....	10
2- Analyse du Prix du Baril au Nigéria de 1960 à 2019.....	11
3- Analyse du PIB par Habitant au Nigéria de 1960 à 2019.....	14
<b>II- ANALYSE DES VARIABLES QUALITATIVES .....</b>	<b>16</b>
1- Analyse de la Balance Commerciale au Nigéria de 1960 à 2019 .....	16
2- Analyse du Solde Budgétaire au Nigéria de 1960 à 2019 .....	17
<b>TROISIEME PARTIE : ANALYSE BIVARIEE.....</b>	<b>18</b>
<b>I- ANALYSE DE DEUX VARIABLES QUANTITATIVES : PB et PIBH .....</b>	<b>18</b>
1- Représentation du nuage de points du <i>Prix du Baril</i> et du <i>PIB par Habitant</i> .....	18
2- Calcul et interprétation de la covariance .....	19
3- Calcul et interprétation du coefficient de corrélation.....	20
4- Calcul et interprétation du coefficient de détermination.....	21
5- Test de significativité de la corrélation .....	21
6- Ajustement linéaire entre le PB et le PIBH .....	22
<b>II- ANALYSE DE DEUX VARIABLES QUANTITATIVES : PB et TXCA.....</b>	<b>22</b>
1- Représentation du nuage de points du Prix de Baril et du Taux de Croissance Annuel .....	22
2- Calcul et interprétation de la covariance .....	23
3- Calcul et interprétation du coefficient de corrélation.....	23
4- Calcul et interprétation du coefficient de détermination .....	23
5- Test de significativité de la corrélation .....	23
<b>III- ANALYSE D'UNE VARIABLE QUANTITATIVE ET D'UNE VARIABLE</b>	
<b>QUALITATIVE : PB et BC .....</b>	<b>24</b>
1- Représentation graphique du Prix du Baril par niveau de facteur de la Balance Commerciale ....	24
2- Calcul du rapport de corrélation et son interprétation.....	24
3- Test de significativité de la corrélation .....	25
<b>CONCLUSION .....</b>	<b>25</b>
<b>ANNEXES .....</b>	<b>26</b>
<b>SOURCE DU CODE R .....</b>	<b>29</b>

## LISTE DES TABLEAUX

Tableau 1 : Dictionnaire de données .....	6
Tableau 2 : Structure du jeu de données.....	6
Tableau 3 : Visualisation des 5 premières observations.....	6
Tableau 4 : Résumés numériques du Prix du Baril .....	13
Tableau 5 : Résumés numériques du PIB par Habitant.....	15
Tableau 6 : Tableau statistique de la Balance Commerciale .....	16
Tableau 7 : Tableau statistique du Solde Budgétaire .....	17

## LISTE DES GRAPHIQUES

Graphique 1 : Visualisation des valeurs manquantes .....	7
Graphique 2 : Autre visualisation des valeurs manquantes .....	7
Graphique 3 : Visualisation des valeurs manquantes après traitement.....	8
Graphique 4 : Visualisation des valeurs aberrantes et extrêmes .....	9
Graphique 5 : Visualisation des valeurs aberrantes et extrêmes après traitement.....	9
Graphique 6 : Courbe d'évolution du Prix du Baril au Nigéria de 1960 à 2019 .....	12
Graphique 7 : Histogramme du PIB par Habitant au Nigéria de 1960 à 2019.....	12
Graphique 8 : Courbe d'évolution du PIB par Habitant au Nigéria de 1960 à 2019 .....	14
Graphique 9 : Histogramme du PIB par Habitant au Nigéria de 1960 à 2019.....	14
Graphique 11 : Diagramme en secteurs de la Balance Commerciale.....	16
Graphique 10 : Diagramme en barres de la Balance Commerciale.....	16
Graphique 12 : Diagramme en barres du Solde Budgétaire .....	17
Graphique 13 : Diagramme en secteurs du Solde Budgétaire .....	17
Graphique 14 : Nuage de points du Prix du Baril et du PIB par Habitant.....	18
Graphique 15 : Nuage de points du Prix du Baril et du Taux de Croissance Annuel .....	22
Graphique 16 : Représentation graphique du Prix du Baril par niveau de facteur de la Balance Commerciale .....	24

## SIGLES ET ABREVIATIONS

BC	: Balance Commerciale
BP	: Balance des paiements
CN	: Cours du Naira
INSEDS	: Institut Supérieur de Statistique d'Econométrie et de Data Science
PB	: Prix du Baril
PIBH	: PIB par Habitant
SB	: Solde Budgétaire
TXCA	: Taux de Croissance Annuel
TXCH	: Taux de Chômage

## INTRODUCTION GENERALE

Le Nigeria, en tant que nation africaine à la riche abondance pétrolière, occupe une place prépondérante sur la scène mondiale en raison de sa participation à l'industrie pétrolière. Depuis son indépendance en 1960, le pays a connu une trajectoire économique influencée de manière significative par les fluctuations du prix du pétrole sur les marchés internationaux. Cette situation a engendré des conséquences majeures pour l'économie nationale, la politique intérieure et les relations internationales du Nigeria.

L'évolution du prix du pétrole au Nigeria entre 1960 et 2019 suscite un intérêt particulier en raison de son impact durable sur divers aspects socio-économiques. Comment les variations des prix du pétrole ont-elles influencé l'économie nigériane au fil des décennies ?

Ce rapport se propose d'explorer l'évolution du prix du pétrole au Nigeria sur la période allant de 1960 à 2019. Pour ce faire, nous allons d'abord prétraiter le jeu de données pour avoir des données saines. Ensuite, nous allons procéder à des analyses univariées et bivariées des variables d'intérêt.

Le présent document s'articulera autour de trois grandes parties :

- PREMIERE PARTIE : PRETRAITEMENT DES DONNEES
- DEUXIEME PARTIE : ANALYSE UNIVARIEE
- TROISIEME PARTIE : ANALYSE BIVARIEE

## PREMIERE PARTIE : PRETRAITEMENT DES DONNEES

Le prétraitement des données avant l'analyse est une étape essentielle pour garantir la qualité et la pertinence des résultats d'analyse. Cette étape consiste à préparer, nettoyer et transformer les données brutes pour qu'elles soient adaptées à une analyse spécifique.

Cette partie consiste, dans un premier temps, à importer le jeu de données dans un logiciel statistique. Ensuite, on procèdera à la visualisation du jeu de données afin d'en dégager la structure. Enfin, la dernière étape consistera à apurer le jeu de données. L'apurement de données, souvent appelé "data cleansing" en anglais, est le processus de nettoyage et de correction des données stockées dans une base de données ou un ensemble de données.

L'objectif de l'apurement de données est d'identifier et de corriger les incohérences, les erreurs, les doublons et les données obsolètes ou incorrectes afin d'assurer la qualité et la précision des données.

### I- DESCRIPTION DU JEU DE DONNEES : dictionnaire des données

Le dictionnaire de données est une documentation détaillée qui répertorie et décrit chacune des variables du jeu de données. Il sert de référence pour comprendre la signification, la structure et les propriétés des données stockées, ce qui facilite la gestion, la maintenance, l'analyse et l'utilisation des données.

Le jeu de données de notre étude se décrit comme suit :

**Tableau 1 : Dictionnaire de données**

VARIABLE	NATURE	DESCRIPTION	MODALITES
ANNEE	Quantitative	Année d'observation	Numérique entier
PB	Quantitative	Prix du Baril	Numérique décimal
BC	Qualitative	Balance Commerciale	Défavorable/favorable
TXCH	Quantitative	Taux de Chômage	Numérique décimal
PIBH	Quantitative	PIB par Habitant	Numérique décimal
CN	Quantitative	Cours du Naira	Numérique décimal
SB	Qualitative	Solde Budgétaire	Excédent / déficit
TXCA	Quantitative	Taux de Croissance Annuel	Numérique décimal
BP	Qualitative	Balance des paiements	Equilibre / déséquilibre

*Source : Informations tirées du site [www.banquemondiale.org](http://www.banquemondiale.org), Nos calculs*

## STRUCTURE DU JEU DE DONNEES

Notre jeu de données est constitué de 60 observations et de 9 variables. On compte 6 variables quantitatives et 3 variables qualitatives à deux facteurs.

**Tableau 2 : Structure du jeu de données**

'data.frame'	60 obs. of 9 variables												
\$ annees	int	1960	1961	1962	1963	1964	1965	1966	1967	1968	1969	...	
\$ PB	num	1.63	1.57	1.52	1.5	1.45	1.42	1.36	1.33	1.32	1.27	...	
\$ BC	Factor w/ 2 levels	"defavorable",... 1 1 1 1 1 1 NA 1 1 1 ...											
\$ TXCH	num	2.43	2.41	2.44	2.47	2.5	2.45	2.37	2.39	2.43	2.51	...	
\$ PIBH	num	93	97	104	108	113							
\$ CN	num	0.714	0.714	0.714	0.714	0.714	0.714	0.714	0.714	0.714	0.714	...	
\$ SB	Factor w/ 2 levels	"deficit","excedent" 2 2 2 2 2 2 2 2 2 ...											
\$ TXCA	num	0.72	0.192	4.103	8.579	4.95							
\$ BP	Factor w/ 2 levels	"desequilibre",... 2 2 2 2 2 2 2 2 2 ...											

*Source : Informations tirées du site [www.banquemondiale.org](http://www.banquemondiale.org), Nos calculs*

## VISUALISATION DU JEU DE DONNEES BRUTES

Le tableau ci-dessous présente les cinq (5) premières observations de notre jeu de données.

**Tableau 3 : Visualisation des 5 premières observations**

N°	annees	PB	BC	TXCH	PIBH	CN	SB	TXCA	BP
1	1960	1.63	defavorable	2.43	92.96	0.714	excedent	0.720	equilibre
2	1961	1.57	defavorable	2.41	96.98	0.714	excedent	0.192	equilibre
3	1962	1.52	defavorable	2.44	104.39	0.714	excedent	4.103	equilibre
4	1963	1.50	defavorable	2.47	107.54	0.714	excedent	8.579	equilibre
5	1964	1.45	defavorable	2.50	113.17	0.714	excedent	4.950	equilibre

*Source : Informations tirées du site [www.banquemondiale.org](http://www.banquemondiale.org), Nos calculs*

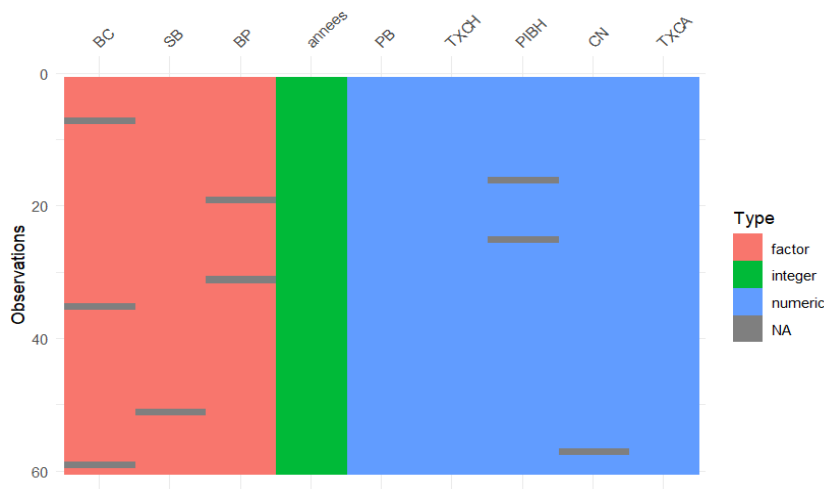
## II- TRAITEMENT DES VALEURS MANQUANTES

En statistique, une valeur manquante, également appelée donnée manquante ou observation manquante, fait référence à l'absence d'une valeur pour une variable particulière dans un ensemble de données ou un échantillon.

### VISUALISATION DES DONNEES MANQUANTES

La visualisation du graphique 1 montre bien que le jeu de données contient des valeurs manquantes. Les variables qui possèdent ces valeurs sont au nombre de 5 dont 2 variables quantitatives et 3 qualitatives.

Graphique 1 : Visualisation des valeurs manquantes

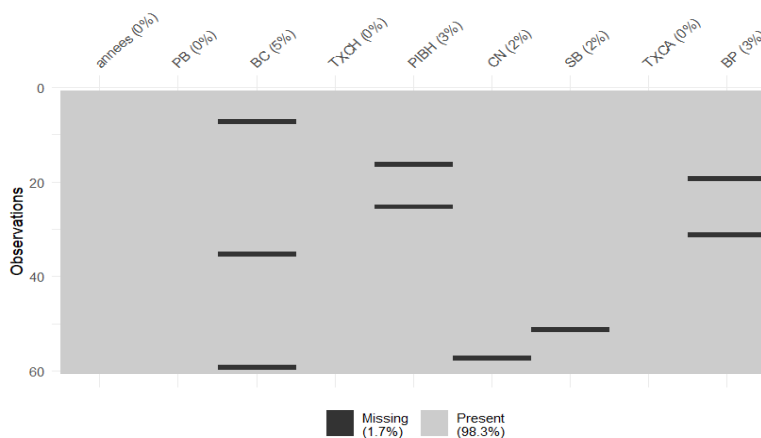


On compte 9 observations avec des valeurs manquantes sur les 60, soit 15% des observations ne possèdent pas de données pour certaines variables (voir graphique 1).

Pour plus de détails sur ces observations (voir annexe 2)

Source : Informations tirées du site [www.banquemondiale.org](http://www.banquemondiale.org), Nos calculs

Graphique 2 : Autre visualisation des valeurs manquantes



Il existe une autre méthode de visualisation des valeurs manquantes (voir graphique 2)

Le jeu de données contient 1,7% de valeurs manquantes.

Source : Informations tirées du site [www.banquemondiale.org](http://www.banquemondiale.org), Nos calculs

Il existe plusieurs méthodes de traitement des valeurs manquantes parmi lesquelles on peut citer :

- supprimer les observations avec des valeurs manquantes ;
- remplir les valeurs manquantes avec les valeurs des caractéristiques de tendance centrale ;
- remplir les valeurs inconnues en explorant les corrélations ;
- remplir les valeurs inconnues en explorant les similitudes entre les cas ;
- remplir les valeurs inconnues par ACP.

Pour notre étude, il est impossible de procéder à une suppression des observations pour traiter les valeurs manquantes, car on a plus de 5% des observations qui possèdent des données inconnues.

Le traitement des valeurs manquantes se fera en deux phases :

➤ Variables quantitatives : les valeurs inconnues sont remplies en explorant les similitudes entre les cas. Cette méthode consiste à trouver les K enregistrements les plus similaires (les plus proches voisins) à l'observation avec des données manquantes, puis à imputer la valeur manquante en fonction des valeurs de ces voisins.

**Pour cette étude, on préfère l'imputation de la médiane à la moyenne car le test de Shapiro montre que ces variables (PIBH, CN) ne suivent pas la loi normale.**

➤ Variables qualitatives : on procède également par imputation pour remplir les valeurs qui sont inconnues (la similitude entre les cas).

*Graphique 3 : Visualisation des valeurs manquantes après traitement*



Source : Informations tirées du site [www.banquemondiale.org](http://www.banquemondiale.org), Nos calculs

Le graphique 3 montre que le jeu de données ne contient plus de valeurs manquantes. Elles ont toutes été traitées avec succès (voir annexe 1 pour plus de détails)

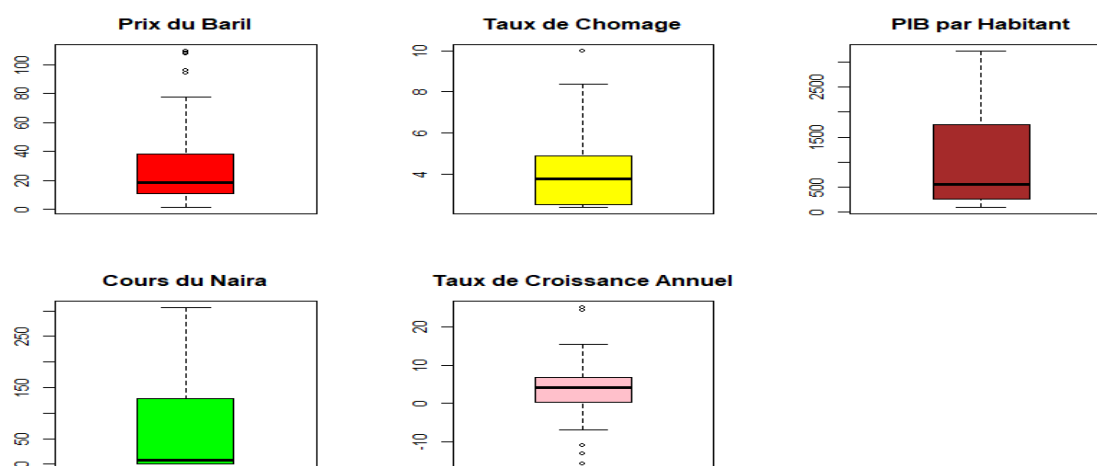


### III- TRAITEMENT DES VALEURS EXTREMES ET ABERRANTES

Les valeurs aberrantes (ou outliers en anglais) et les valeurs extrêmes (ou extreme values) sont des observations dans un ensemble de données qui se situent à l'extérieur de la plage attendue ou qui diffèrent considérablement du reste des données. Elles peuvent avoir un impact significatif sur les analyses statistiques et les modèles, car elles peuvent influencer les résultats de manière disproportionnée.

Les valeurs aberrantes et extrêmes concernent uniquement les variables quantitatives. La visualisation de ces valeurs se fait à l'aide des boîtes à moustache. Une valeur est considérée comme aberrante ou extrême lorsqu'elle est située au-delà des moustaches.

*Graphique 4 : Visualisation des valeurs aberrantes et extrêmes*



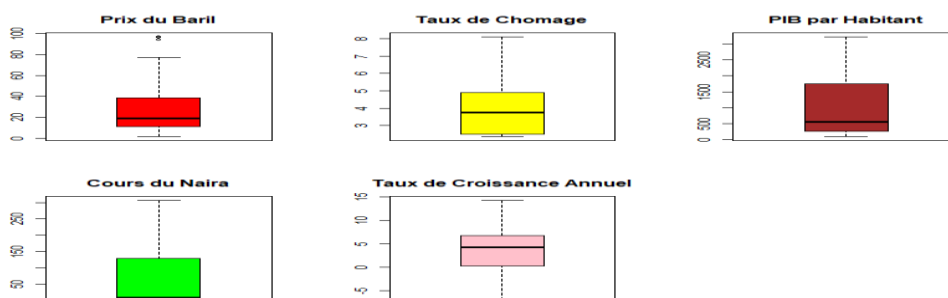
Source : Informations tirées du site [www.banquemondiale.org](http://www.banquemondiale.org), Nos calculs

On voit bien sur le graphique 4 que les variables *Prix de Baril*, *Taux de Chômage* et *Taux de croissance Annuel* contiennent des valeurs aberrantes et extrêmes.

Pour traiter les valeurs aberrantes observées sur le graphique, nous allons utiliser la winsorisation. C'est une technique de traitement des valeurs aberrantes dans un jeu de données qui consiste à ramener ces valeurs dans la limite des bornes (inférieure et supérieure). Cette méthode doit son nom à son concepteur, Charles P. Winsor, un statisticien américain.

Après avoir traité les valeurs aberrantes, la visualisation nous donne ci-dessous (graphique 5) :

*Graphique 5 : Visualisation des valeurs aberrantes et extrêmes après traitement*



Source : Informations tirées du site [www.banquemondiale.org](http://www.banquemondiale.org), Nos calculs

## DEUXIEME PARTIE : ANALYSE UNIVARIEE

Dans cette partie du document, il s'agira pour nous de définir et interpréter les indicateurs de tendance centrale, les indicateurs de dispersion et les indicateurs de forme.

### I- ANALYSE DES VARIABLES QUANTITATIVES

#### 1- Rappel du fondement théorique des indicateurs statistiques

##### a. Les indicateurs de tendance centrale ou de position

Les indicateurs de tendance centrale sont des mesures statistiques qui visent à résumer ou à représenter le centre ou la valeur centrale d'un ensemble de données. Les indicateurs de tendance centrale couramment utilisés sont les suivants :

- **MINIMUM** : la plus petite valeur des observations
- **MAXIMUM** : la plus grande valeur des observations
- **MODE** : Le mode est la valeur de la variable statistique qui a le plus grand effectif ou la plus grande fréquence (il peut y avoir plusieurs modes)
- **MEDIANE** : La médiane d'une série statistique est le nombre qui partage cette série en deux séries de même effectif. La moitié des effectifs (50 %) a donc une valeur du caractère en dessous de la valeur médiane et l'autre moitié (50 %) au-dessus.

**Q1** : Le quartile Q1, la médiane Me et le quartile Q3 partagent les valeurs ordonnées de la série en quatre parties égales. Q1 est la plus petite donnée de la série pour laquelle au moins 25 % des données (soit 1/4 des données) sont égales ou inférieures à Q1 (Q1 est obligatoirement une donnée de la série).

**Q2** : Q2 correspond à la médiane

**Q3** : Q3 est la plus petite donnée de la série pour laquelle au moins 75 % des données (soit 3/4 des données) sont égales ou inférieures à Q3 (Q3 est obligatoirement une donnée de la série).

- **MOYENNE** : c'est la valeur moyenne de la série statistique

##### b. Les indicateurs de dispersion

Les indicateurs de dispersion, également appelés mesures de variabilité, sont des mesures statistiques qui permettent de quantifier la dispersion ou la répartition des données dans un ensemble de données. Ils donnent des informations sur la façon dont les valeurs s'écartent de la tendance centrale (comme la moyenne, la médiane ou le mode) et aident à comprendre la dispersion des données. Voici quelques indicateurs de dispersion couramment utilisés :

- **VARIANCE** : c'est la moyenne des carrés des écarts à la moyenne
- **ECART-TYPE** : L'écart-type traduit une moyenne d'écarts autour de la valeur moyenne. C'est la racine carrée de la variance. Plus l'écart-type est grand, plus la dispersion des données est grande.
- **COEFFICIENT DE VARIATION** : Généralement exprimé en %, Il est notamment utile dans le cas où l'on souhaite comparer deux groupes d'observations.

Un *cv* faible indique une distribution homogène

Un *cv* élevé indique une distribution hétérogène

### c. Les indicateurs de forme

Les indicateurs de forme, également appelés mesures de forme, sont des mesures statistiques utilisées pour caractériser la distribution ou la forme d'un ensemble de données. Voici quelques indicateurs de forme couramment utilisés :

- **SKEWNESS** : L'asymétrie mesure la dissymétrie ou l'absence de symétrie dans la distribution des données par rapport à la moyenne. Elle indique la direction et le degré d'inclinaison de la distribution.

*Une asymétrie positive (skewness positive) indique que la queue droite de la distribution est plus longue que la queue gauche. La plupart des valeurs se trouvent à gauche de la moyenne.*

*Une asymétrie négative (skewness négative) indique que la queue gauche de la distribution est plus longue que la queue droite. La plupart des valeurs se trouvent à droite de la moyenne.*

*Une asymétrie proche de zéro suggère une distribution approximativement symétrique.*

- **KURTOSIS** : Le kurtosis mesure l'aplatissement ou l'effilochage de la distribution par rapport à une distribution normale (gaussienne).

*Une kurtosis élevée indique une distribution avec des valeurs plus concentrées autour de la moyenne (plus pointue) et des queues plus épaisses*

*Une kurtosis faible indique une distribution plus aplatie avec des queues moins épaisses que la distribution normale.*

### 2- Analyse du Prix du Baril au Nigéria de 1960 à 2019

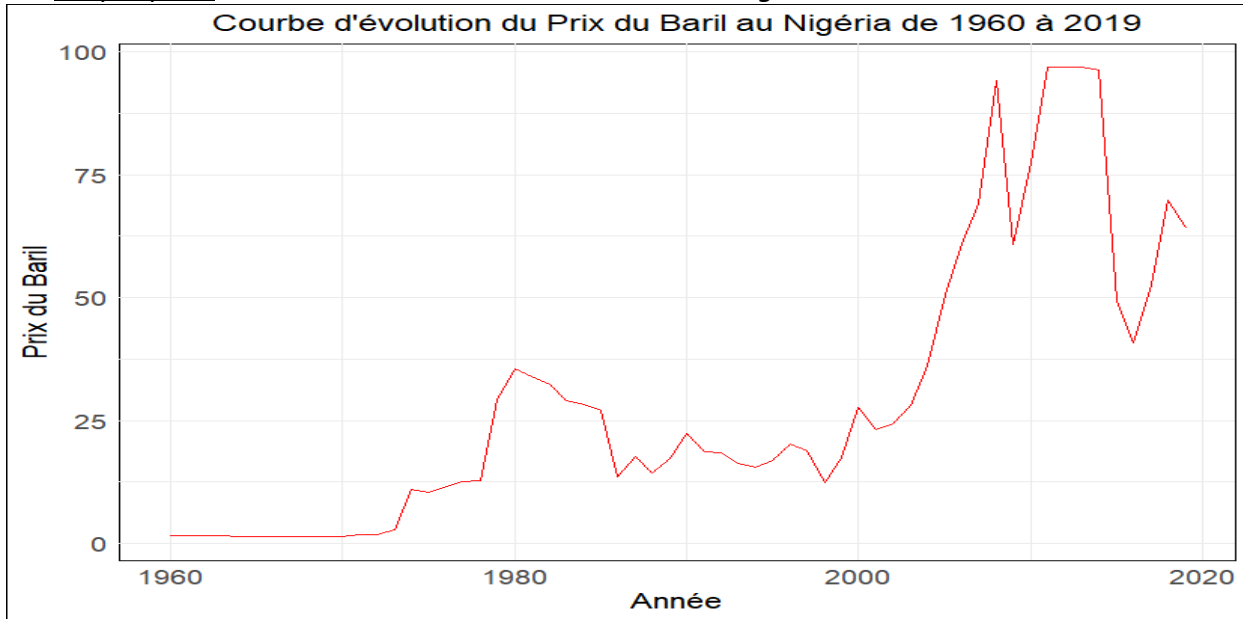
La variable *Prix du Baril* étant continue avec plusieurs valeurs isolées par année (série temporelle), nous allons nous contenter de la représenter par une courbe d'évolution qui permet de mieux à analyser une telle variable.

De manière générale, on constate que le *Prix du Baril* a évolué au Nigéria de 1960 à 2019 avec plusieurs pics comme le montre bien le graphique 6. Cela traduit l'instabilité du prix du pétrole au cours de cette période d'étude. De 1960 à 2000, on constate une croissance lente du *Prix du Baril* tandis qu'à partir de 2000, le *Prix du Baril* a connu une forte croissance en présentant des situations qui n'avaient jamais été observées auparavant.

De manière spécifique, on remarque que le *Prix du Baril* est presque constant de 1960 à 1973. En outre, on note que des pics ont été atteints en 1980 (35,52 Nairas), 1990 (22,26 Nairas), 2000 (27,6 Nairas) et 2008 (94,1 Nairas, prix que le pays n'avait jamais connu depuis l'exploitation du pétrole).

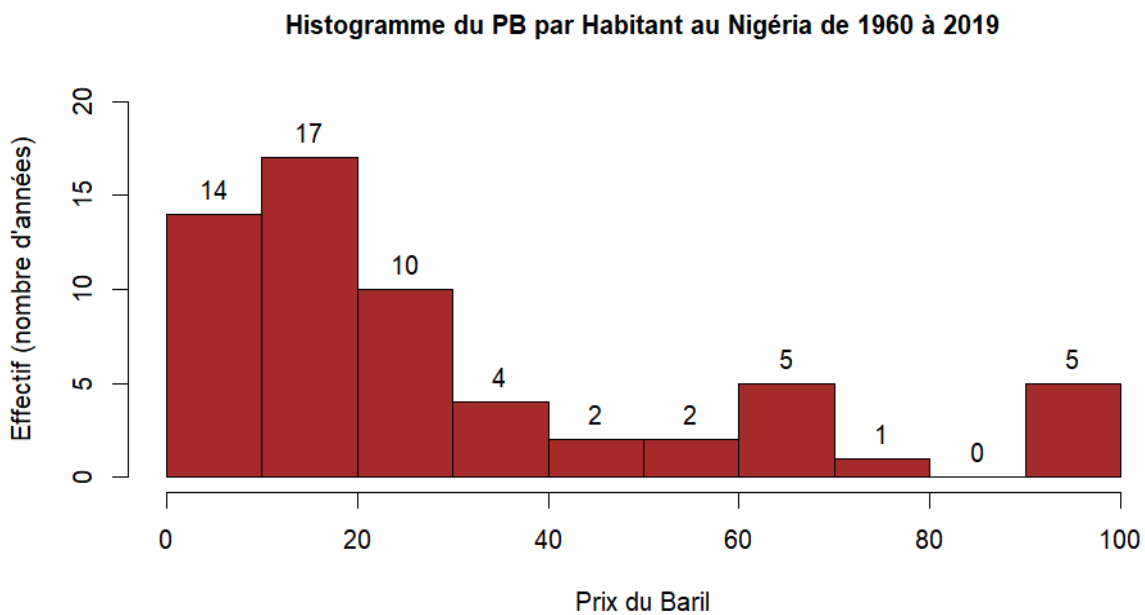
L'année 2011 a été particulière dans la mesure où le prix du pétrole a connu une croissance pour atteindre son maximum (96,8485 Nairas) et a stagné jusqu'en 2013.

**Graphique 6 : Courbe d'évolution du Prix du Baril au Nigéria de 1960 à 2019**



Source : Informations tirées du site [www.banquemondiale.org](http://www.banquemondiale.org), Nos calculs

**Graphique 7 : Histogramme du PIB par Habitant au Nigéria de 1960 à 2019**



Source : Informations tirées du site [www.banquemondiale.org](http://www.banquemondiale.org), Nos calculs

Le tableau 4 présente les résumés numériques du *Prix du Baril* (PB)

*Tableau 4 : Résumés numériques du Prix du Baril*

INDICATEUR	VALEUR	INTERPRETATION
<b>INDICATEURS DE TENDANCE CENTRALE ET DE POSITION</b>		
Minimum	1,33	Le prix minimum du pétrole est 1,33 Naira
Maximum	96,85	Le prix maximum du pétrole est 96,85 Nairas
Mode	14,68	On a pratiqué un prix égal 14,68 Nairas sur la majorité des années.
Moyenne	29,18	Le prix moyen du pétrole est 29,18 Nairas
1 <sup>er</sup> quartile Q1	10,86	25% des années ont connu un prix maximum de 10,86 Nairas
Médiane (2 <sup>eme</sup> quartile Q2)	18,74	La moitié des années a connu un prix maximum de 18,74 Nairas
3 <sup>eme</sup> quartile Q3	37,23	75% des années ont connu un prix maximum de 37,23 Nairas
<b>INDICATEURS DE DISPERSION</b>		
Variance	802,12	On interprétera l'écart-type qui est la racine carrée de la variance
Ecart-type	28,32	Le prix moyen étant 29,18 Nairas. Un écart-type de 28,32 indique que les prix sont dispersés et en moyenne ils varient dans cet intervalle [29,18 - 28,32 ; 29,18 + 28,32] soit [0,86 ; 57,18]
Coefficient de variation	97,05	Le coefficient de variation est de 97,05%, ce qui indique que la distribution des prix de pétrole est très hétérogène
<b>INDICATEURS DE FORME</b>		
Skewness (Asymétrie)	1,16	Le Skewness étant positif, cela indique que la distribution des prix de pétrole est étalée à droite. (Voir graphique 7)
Kurtosis (Aplatissement)	3,34	Le Kurtosis étant supérieur à 3, la distribution des prix du baril est dite leptokurtique, c'est-à-dire qu'elle est plus aigüe que celle d'une loi normale. (Voir graphique 7)

*Source : Informations tirées du site [www.banquemondiale.org](http://www.banquemondiale.org), Nos calculs*

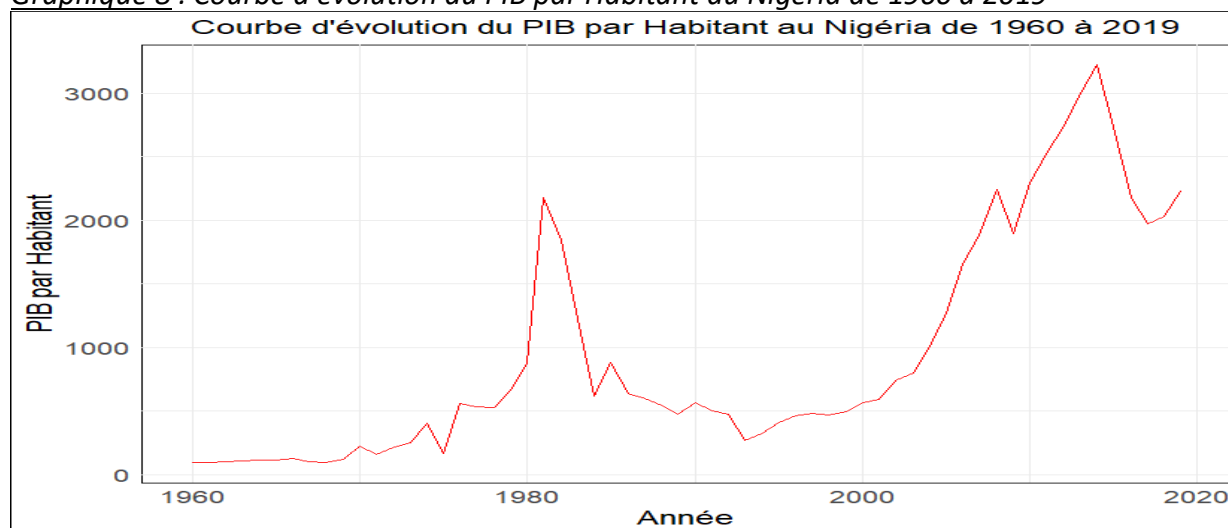
### 3- Analyse du PIB par Habitant au Nigéria de 1960 à 2019

A l'image du *Prix du Baril* précédemment étudié, le *PIB par Habitant* est aussi une série temporelle. On peut alors le représenter par une courbe d'évolution.

On remarque que de manière générale le *PIB par Habitant* a connu une croissance de 1960 à 2019 avec plusieurs pics, cela signifie que le *PIB par Habitant* a été instable sur la période.

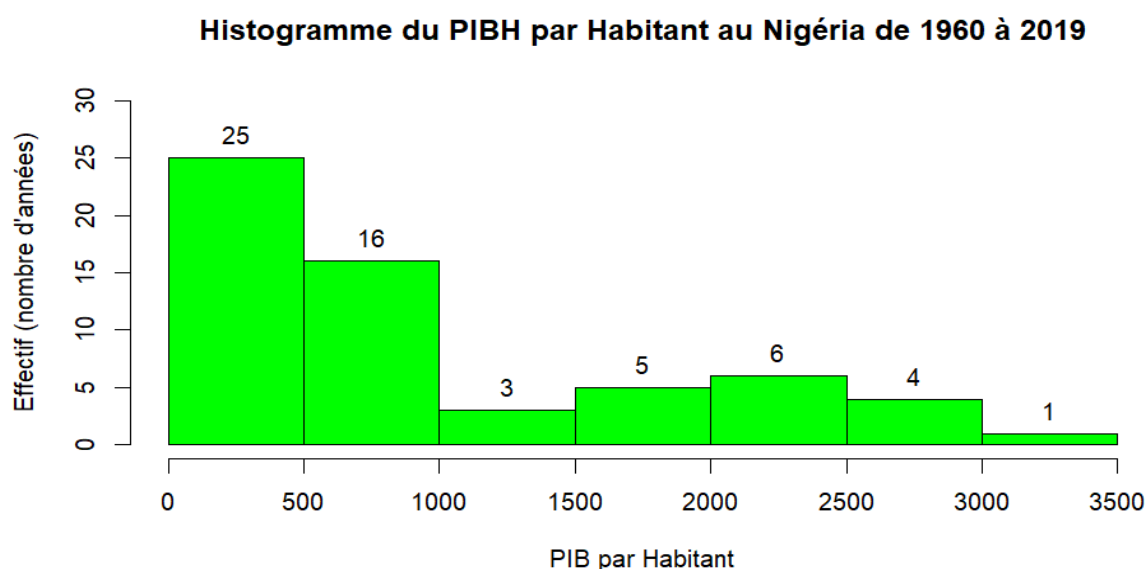
De 1960 à 1994, on observe une faible croissance de la variable étudiée avec un pic assez pointu en 1981. Mais à partir de 1994, la croissance du *PIB par Habitant* a accéléré pour atteindre sa valeur maximale qui est 3222,69 en 2014 avant de baisser à nouveau.

**Graphique 8 : Courbe d'évolution du PIB par Habitant au Nigéria de 1960 à 2019**



Source : Informations tirées du site [www.banquemondiale.org](http://www.banquemondiale.org), Nos calculs

**Graphique 9 : Histogramme du PIB par Habitant au Nigéria de 1960 à 2019**



Source : Informations tirées du site [www.banquemondiale.org](http://www.banquemondiale.org), Nos calculs

Le tableau 5 présente les résumés numériques du *PIB par Habitant*.

**Tableau 5 : Résumés numériques du PIB par Habitant**

INDICATEUR	VALEUR	INTERPRETATION
<b>INDICATEURS DE TENDANCE CENTRALE ET DE POSITION</b>		
Minimum	92,96	La valeur minimale du PIBH est de 92,96
Maximum	3222,69	La valeur maximale du PIBH est de 3222,69
Mode	412,38	La majorité des années a connu une valeur du PIBH de 412,38
Moyenne	944,51	Le PIBH moyen est de 944,51
1 <sup>er</sup> quartile Q1	265,72	25% des années ont connu un PIBH maximum de 265,72
Médiane (2 <sup>ème</sup> quartile Q2)	561,97	La moitié des années a connu un PIBH maximum de 561,97
3 <sup>ème</sup> quartile Q3	1703,29	75% des années a connu un PIBH maximum de 1703,29
<b>INDICATEURS DE DISPERSION</b>		
Variance	784040,10	On interprétera l'écart-type qui est la racine carrée de la variance
Ecart-type	885,46	Le PIBH moyen est de 944,51. Un écart-type de 885,46 indique que les PIBH sont dispersés et en moyenne ils varient dans cet intervalle [944,51 - 885,46 ; 944,51 + 885,46] soit [59,05 ; 1829,51]
Coefficient de variation	93,75	Le coefficient de variation est de 93,75%, ce qui indique que la distribution des prix de pétrole est très hétérogène
<b>INDICATEURS DE FORME</b>		
Skewness (Asymétrie)	1,02	Le Skewness étant positif, cela indique que la distribution des prix de pétrole est étalée à droite. (Voir graphique 9)
Kurtosis (Aplatissement)	2,74	Le Kurtosis étant inférieur à 3, cela indique la distribution des prix de baril est moins aigüe que celle d'une loi normale. (Voir graphique 9)

Source : Informations tirées du site [www.banquemondiale.org](http://www.banquemondiale.org), Nos calculs

## II- ANALYSE DES VARIABLES QUALITATIVES

### 1- Analyse de la Balance Commerciale au Nigéria de 1960 à 2019

#### a. Tableau statistique de la *Balance Commerciale*

Tableau 6 : Tableau statistique de la Balance Commerciale

Balance commerciale	Effectif (nombre d'années)	Fréquence
defavorable	24	40%
favorable	36	60%
Total	60	100%

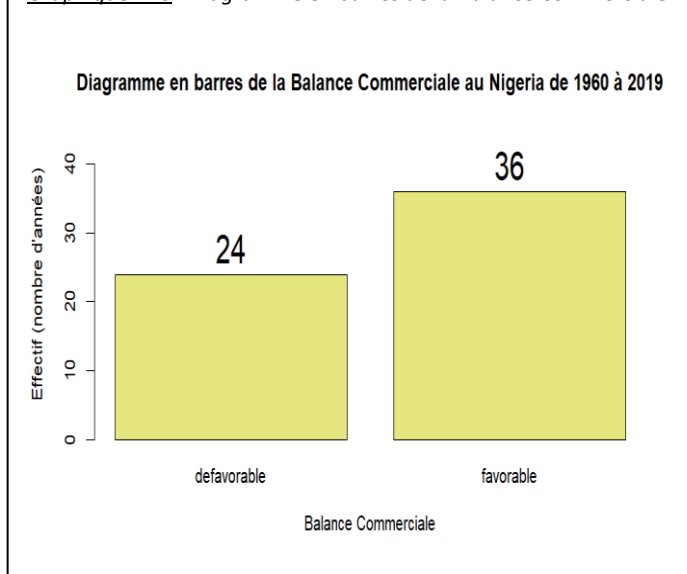
Source : Informations tirées du site [www.banquemondiale.org](http://www.banquemondiale.org), Nos calculs

Le tableau ci-contre montre qu'au cours des 60 années d'étude, le Nigéria a connu une balance commerciale favorable pour 36 années, soit 60% des années allant de 1960 à 2019 ont affiché une balance commerciale favorable.

#### b. Diagramme en barres et diagramme en secteurs de la *Balance Commerciale*

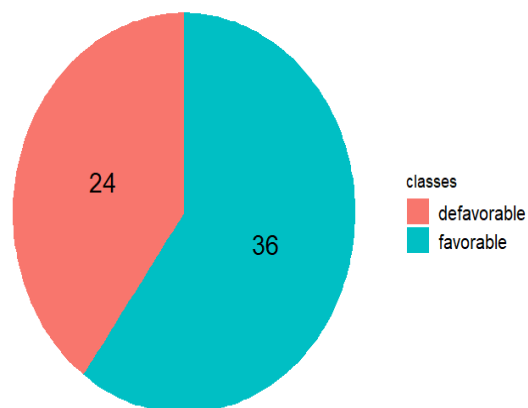
Les graphiques 10 et 11 illustrent la situation de la balance commerciale nigériane décrite dans le paragraphe précédent.

Graphique 110 : Diagramme en barres de la Balance Commerciale



Graphique 101 : Diagramme en secteurs de la Balance Commerciale

Diagramme en secteurs de la Balance Commerciale au Nigéria de 1960 à 2019



Source : Informations tirées du site [www.banquemondiale.org](http://www.banquemondiale.org), Nos calculs



## 2- Analyse du Solde Budgétaire au Nigéria de 1960 à 2019

### a. Tableau statistique du *Solde Budgétaire*

**Tableau 7 : Tableau statistique du Solde Budgétaire**

Solde budgétaire	Effectif (nombre d'années)	Frequence
deficit	22	37%
excedent	38	63%
Total	60	100%

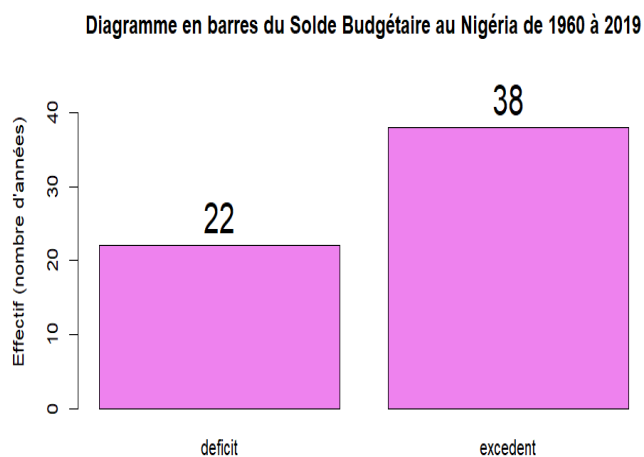
Source : Informations tirées du site [www.banquemondiale.org](http://www.banquemondiale.org), Nos calculs

Le tableau ci-contre indique que sur les 60 années d'étude (1960 à 2019), 38 ont enregistré un Solde budgétaire excédentaire au Nigéria.

### b. Diagramme en barres et diagramme en secteurs du *Solde Budgétaire*

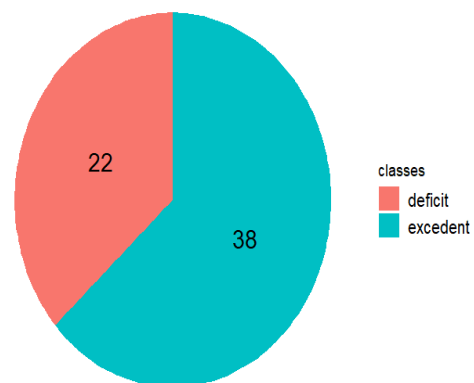
Nous percevons mieux la situation décrite dans le paragraphe précédent sur les graphiques 12 et 13.

**Graphique 12 : Diagramme en barres du Solde Budgétaire**



**Graphique 13 : Diagramme en secteurs du Solde Budgétaire**

Diagramme en secteurs du Solde Budgétaire au Nigéria de 1960 à 2019



Source : Informations tirées du site [www.banquemondiale.org](http://www.banquemondiale.org), Nos calculs

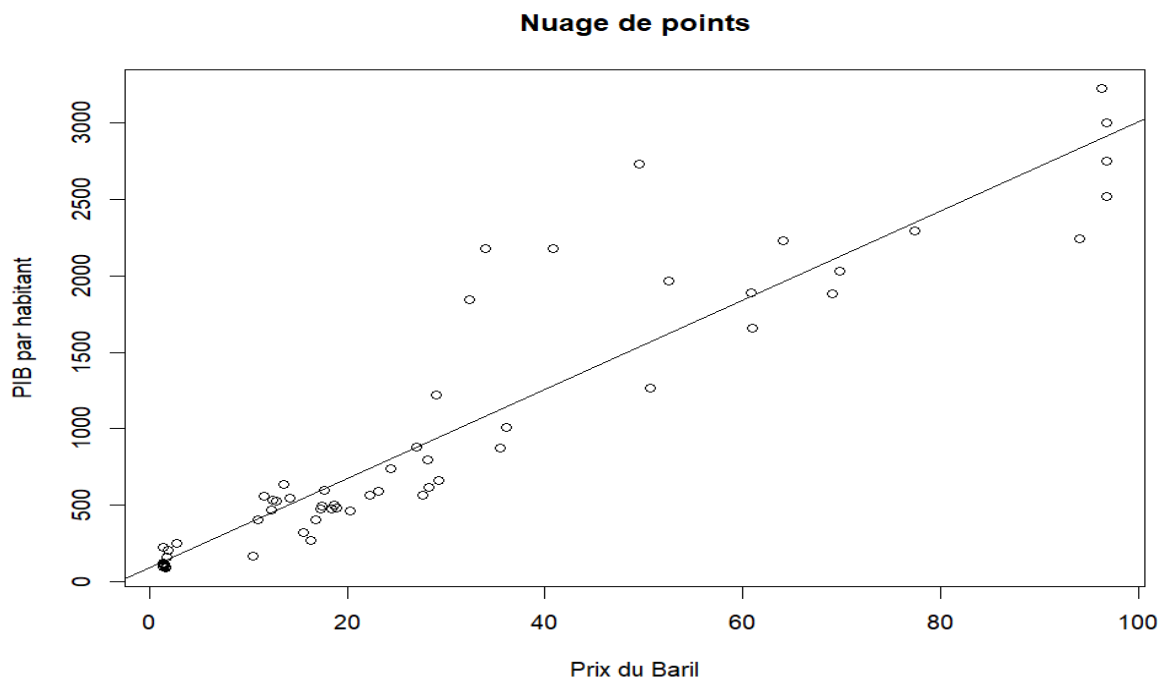
## TROISIEME PARTIE : ANALYSE BIVARIEE

### I- ANALYSE DE DEUX VARIABLES QUANTITATIVES : PB et PIBH

#### 1- Représentation du nuage de points du *Prix du Baril* et du *PIB par Habitant*

Le graphique ci-dessous nous montre que le nuage de points a l'allure d'une droite, on peut alors soupçonner l'existence d'un lien entre le *Prix du Baril* et le *PIB par Habitant*. Pour lever ce doute, on va procéder au calcul du coefficient de détermination en passant par la covariance et le coefficient de détermination.

Graphique 14 : Nuage de points du *Prix du Baril* et du *PIB par Habitant*



Source : Informations tirées du site [www.banquemondiale.org](http://www.banquemondiale.org), Nos calculs

## 2- Calcul et interprétation de la covariance

La covariance est une mesure statistique qui évalue la relation linéaire entre deux ensembles de données. Elle quantifie comment les variations dans un ensemble de données correspondent aux variations dans un autre ensemble de données. La covariance est couramment utilisée pour déterminer si deux variables sont associées positivement, négativement ou ne sont pas du tout associées.

La formule de la covariance entre deux ensembles de données,  $x$  et  $y$ , est la suivante :

$$\begin{aligned} \text{Cov}(x; y) = \sigma_{xy} &= \frac{1}{n} \sum (x_i - \bar{x})(y_i - \bar{y}) \\ \text{ou} \\ \text{Cov}(x; y) = \sigma_{xy} &= \frac{1}{n} \sum x_i y_i - \bar{x} \bar{y} \end{aligned}$$

Où :

$x_i$  et  $y_i$  sont les valeurs individuelles dans les ensembles de données  **$x$  et  $y$**

$\bar{x}$  et  $\bar{y}$  représentent respectivement les moyennes de  **$x$  et  $y$**

$n$  est le nombre total d'observations dans les ensembles de données

### Interprétation de la covariance :

- ❖ *Covariance positive : Si la covariance est positive, cela signifie que lorsque les valeurs de  $X$  augmentent, les valeurs de  $Y$  ont tendance à augmenter, et lorsque les valeurs de  $X$  diminuent, les valeurs de  $Y$  ont tendance à diminuer.*
- ❖ *Covariance négative : Si la covariance est négative, cela signifie que lorsque les valeurs de  $X$  augmentent, les valeurs de  $Y$  ont tendance à diminuer, et vice versa.*
- ❖ *Covariance proche de zéro : Si la covariance est proche de zéro, cela indique qu'il y a peu ou pas de relation linéaire entre les ensembles de données  $X$  et  $Y$ .*

Dans le cas de notre étude, la covariance des variables PB et PIBH vaut 23370,57. La covariance étant positive, on se trouve dans le 1er cas ; c'est-à-dire lorsque le prix du baril augmente, le PIB par habitant augmente et inversement.

**Remarque :** La principale limitation de la covariance est qu'elle n'est pas normalisée, ce qui signifie que sa valeur dépend de l'échelle des données. Par conséquent, la covariance ne permet pas de quantifier la force de la relation entre les variables de manière indépendante de leur échelle. Pour remédier à cela, on utilise souvent le coefficient de corrélation, tel que le coefficient de corrélation de Pearson, qui est une version normalisée de la covariance. Le coefficient de corrélation varie de -1 à 1 et permet de mesurer la force et la direction de la relation linéaire entre deux variables tout en étant indépendant de l'échelle des données.

### 3- Calcul et interprétation du coefficient de corrélation

Le coefficient de corrélation est une mesure statistique qui évalue la force et la direction de la relation linéaire entre deux variables. Il est couramment utilisé pour déterminer si deux variables sont associées de manière positive, négative ou neutre. Le coefficient de corrélation quantifie à quel point les variations dans une variable sont liées aux variations dans une autre variable. L'un des coefficients de corrélation les plus couramment utilisés est le coefficient de corrélation de Pearson, souvent noté  $r$  ou  $\rho$  (rho).

Le coefficient de corrélation de Pearson mesure la relation linéaire entre deux variables continues. Sa formule est la suivante :

$$r = \frac{Cov(x; y)}{\sigma_x \sigma_y} = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$$

$Cov(x; y) = \sigma_{xy}$  représente la covariance des variables  $x$  et  $y$

$\sigma_x$  et  $\sigma_y$  représentent respectivement les écarts-types de  $x$  et  $y$

#### Interprétation du coefficient de corrélation de Pearson :

- ❖  $r = 1$  : Une corrélation positive parfaite. Les valeurs de  $X$  augmentent linéairement lorsque les valeurs de  $Y$  augmentent.
- ❖  $r = -1$  : Une corrélation négative parfaite. Les valeurs de  $X$  diminuent linéairement lorsque les valeurs de  $Y$  augmentent.
- ❖  $r = 0$  : Aucune corrélation linéaire. Les variables  $X$  et  $Y$  ne montrent aucune relation linéaire.

La valeur de  $r$  se situe généralement entre  $-1$  et  $1$ . Plus la valeur absolue de  $r$  est proche de  $1$ , plus la relation linéaire entre les deux variables est forte. Plus la valeur de  $r$  se rapproche de  $0$ , plus la relation linéaire entre les variables est faible.

Dans le cas de notre étude, le coefficient de corrélation de Pearson des variables PB et PIBH vaut  $0,9261172$ . On constate que cette valeur est très proche de  $1$ , cela signifie qu'il existe une liaison très forte entre le *Prix du Baril* et le *PIB par Habitant*.

#### 4- Calcul et interprétation du coefficient de détermination

Le coefficient de détermination, souvent désigné par  $R^2$ , est une mesure statistique utilisée pour évaluer la proportion de la variance totale dans une variable dépendante (ou cible) qui peut être expliquée par les variables indépendantes (ou explicatives) dans un modèle de régression. En d'autres termes, il mesure la qualité de l'ajustement d'un modèle aux données. Plus précisément, le coefficient de détermination indique à quel point les variations dans la variable dépendante peuvent être prédites ou expliquées par le modèle.

Le coefficient de détermination  $R^2$  est généralement exprimé en pourcentage, ce qui signifie qu'il varie de 0 % à 100 %. Voici comment il est calculé :

$$R^2 = r^2$$

avec  $r$  : coefficient de corrélation

#### Interprétation du coefficient de détermination $R^2$ :

- ❖  $R^2 = 0$  : *Aucune variation de la variable dépendante n'est expliquée par le modèle. Le modèle n'est pas utile pour prédire ou expliquer la variable dépendante.*
- ❖  $0 < R^2 < 1$  : *Une partie de la variation de la variable dépendante est expliquée par le modèle. Plus  $R^2$  est proche de 1, plus le modèle est capable d'expliquer la variation*
- ❖  $R^2 = 1$  : *Le modèle explique parfaitement toute la variation de la variable dépendante. Cependant, cela est rare dans la pratique.*

Dans notre étude, le coefficient de détermination des variables étudiées vaut 0,8576931. On peut donc dire que la variabilité du PIBH est expliquée à 85,77% par la variabilité du *Prix du Baril*.

#### 5- Test de significativité de la corrélation

Dans cette partie, on va réaliser un test de significativité pour voir si nos deux variables sont significativement liées comme nous l'avons mentionné plus haut. Ce test nécessite la formulation des hypothèses suivantes :

$H_0: r_{xy} = 0$ , c'est-à-dire les deux variables ne sont pas liées

$H_1: r_{xy} \neq 0$ , c'est-à-dire les deux variables sont liées

La p-value de ce test donne p-value = 3,119234e-26. On voit bien que notre p-value est inférieur à 5%, on ne peut que rejeter l'hypothèse  $H_0$ . Cela voudrait dire qu'on l'hypothèse selon laquelle le PB et le PIBH sont significativement liés.

## 6- Ajustement linéaire entre le PB et le PIBH

Au paragraphe précédent, nous avons dit que le PB et le PIBH sont significativement liés, cela sous-entend qu'on peut rechercher un ajustement linéaire entre les deux variables. En d'autres termes, on peut rechercher la meilleure droite qui résume la structure du nuage des points. En utilisant la méthode des moindres carrés ordinaires, on obtient l'équation de droite suivante :

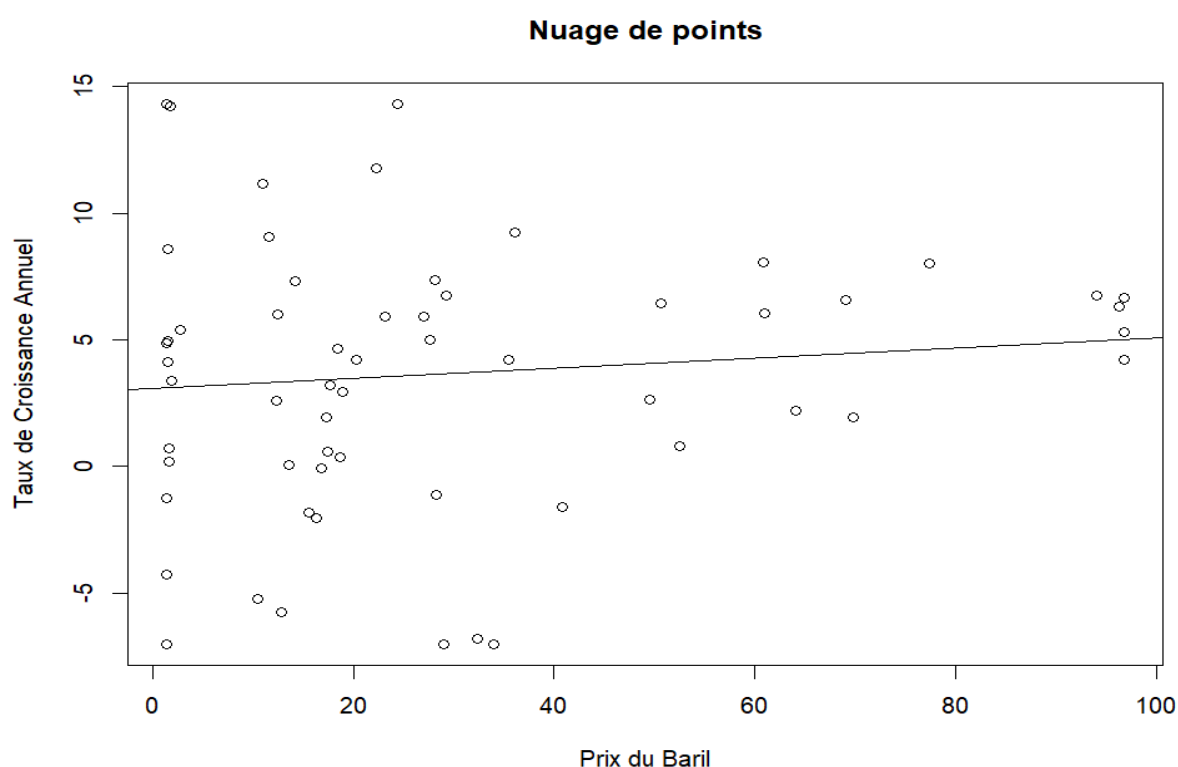
$$\text{PIBH} = 0.03115189 * \text{PB} + 0.32690008$$

## II- ANALYSE DE DEUX VARIABLES QUANTITATIVES : PB et TXCA

### 1- Représentation du nuage de points du Prix de Baril et du Taux de Croissance Annuel

Le graphique montre l'existence d'un faible lien entre les deux variables considérées. Les deux variables étant quantitatives, on peut aller plus loin en calculant le coefficient de détermination qui permettra de corroborer ou infirmer la prédiction du graphique. Mais bien avant, regardons d'abord la covariance et le coefficient de corrélation dans les paragraphes suivants.

Graphique 15 : Nuage de points du Prix du Baril et du Taux de Croissance Annuel



Source : Informations tirées du site [www.banquemondiale.org](http://www.banquemondiale.org), Nos calculs

## 2- Calcul et interprétation de la covariance

La valeur de la covariance des variables *Prix du Baril* (PB) et *Taux de Chômage* (TXCH) est 15,73032. Cette valeur étant positive, on peut dire que lorsque le *Prix du Baril* augmente, le *Taux de Chômage* augmente et inversement.

Nous allons maintenant calculer le coefficient de corrélation de Pearson pour mesurer la relation linéaire entre deux variables.

## 3- Calcul et interprétation du coefficient de corrélation

Dans notre étude, le coefficient de corrélation entre les deux variables vaut 0.1051834. Cette valeur étant très proche de 0, on peut déduire qu'il existe une très faible liaison entre le *Prix du Baril* et le *Taux de Chômage*.

## 4- Calcul et interprétation du coefficient de détermination

Le coefficient de détermination des variables PB et TXCH est 0,01106356. Cela signifie que la variabilité du *Taux de Chômage* est expliquée à 1,11% par la variabilité du *Prix de Baril*.

## 5- Test de significativité de la corrélation

Le test de la corrélation consiste à vérifier si les deux variables sont significativement liées en formulant les hypothèses suivantes :

$H_0: r_{xy} = 0$ , c'est-à-dire les deux variables ne sont pas liées

$H_1: r_{xy} \neq 0$ , c'est-à-dire les deux variables sont liées

La p-value de notre test est : p-value = 0,4238. La p-value dans notre cas est supérieur à 5%. On accepte alors l'hypothèse  $H_0$ , c'est-à-dire que le *Prix du Baril* et le *Taux de Chômage* ne sont pas significativement liées.

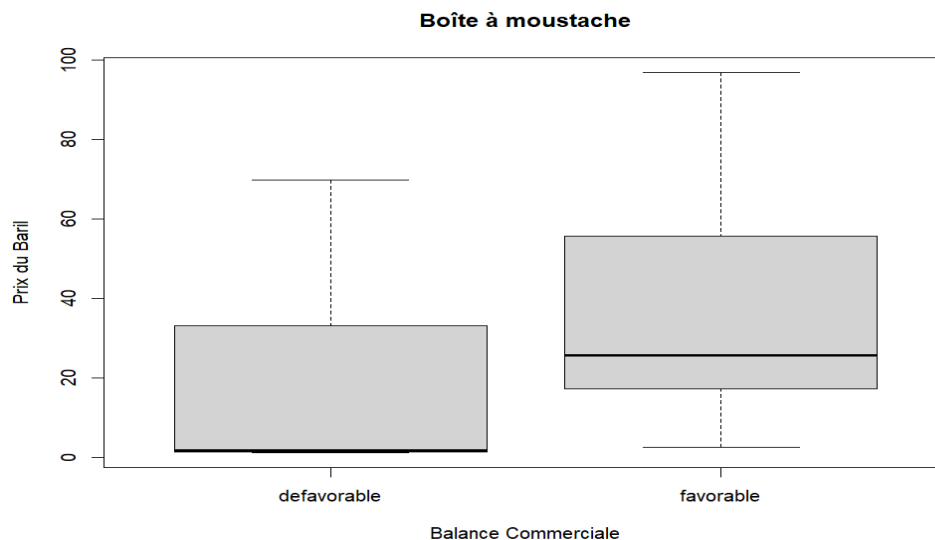
On ne peut pas par conséquent trouver une droite qui résume la structure du nuage des points.

### III- ANALYSE D'UNE VARIABLE QUANTITATIVE ET D'UNE VARIABLE QUALITATIVE : PB et BC

#### 1- Représentation graphique du Prix du Baril par niveau de facteur de la Balance Commerciale

Quand on observe le graphique, on peut soupçonner qu'il y a un lien entre le Prix du Baril et la Balance Commerciale, car les boîtes à moustache ne sont pas parfaitement alignées.

Graphique 16 : Représentation graphique du Prix du Baril par niveau de facteur de la Balance Commerciale



Source : Informations tirées du site [www.banquemondiale.org](http://www.banquemondiale.org), Nos calculs

#### 2- Calcul du rapport de corrélation et son interprétation

Le rapport de corrélation est un indicateur qui permet de vérifier l'existence d'un lien entre une variable quantitative et une variable qualitative. Le rapport de corrélation de Y par rapport à X se définit par :

$$\eta^2_{Y|X} = \frac{V_{inter}}{V_{totale}} = \frac{V_{inter}}{V_{inter} + V_{intra}}$$

- $V_{inter}(Y)$ , variance intergroupe, c'est-à-dire la variance des moyennes de Y par rapport à chaque facteur de X
- $V_{intra}(Y)$ , variance intragroupe, c'est-à-dire la moyenne des variances de Y par rapport à chaque facteur de X
- $V_{totale} = V_{inter}(Y) + V_{intra}(Y)$

Dans le cas de notre étude, ce rapport vaut 0,1242982. On peut donc dire que les deux variables *Prix du Baril* et *Balance Commerciale* sont faiblement liées car leur rapport de corrélation est proche de 0.



### 3- Test de significativité de la corrélation

Ici nous sommes en présence d'une variable quantitative et d'une variable qualitative. Nous allons donc utiliser le test d'ANOVA pour tester la significativité de la corrélation entre le *Prix du Baril* et la *Balance Commerciale*.

La p-value de ce test nous donne  $p\text{-value} = 0,005731063$  qui est inférieur à 5%. On rejette alors l'hypothèse nulle et on accepte l'hypothèse alternative, cela signifie que le prix du baril et la balance commerciale sont significativement liés.

## CONCLUSION

Au terme de notre analyse, nous retenons que l'évolution du prix du baril de pétrole au Nigeria de 1960 à 2019 a été marquée par des hauts et des bas significatifs avec une moyenne qui varie dans l'intervalle  $[0,86 ; 57,18]$ . La moitié des années étudiées a connu un prix maximum du pétrole qui est de 18,74 Nairas.

Le *PIB par Habitant* a aussi connu plusieurs instabilités au cours de la période étudiée. Avec une moyenne de 944,51, il a varié entre 59,05 et 1829,51 de 1960 à 2019.

La balance commerciale du Nigeria a été favorable pour 36 années de 1960 à 2019, soit 60%. Le solde budgétaire, quant à lui, a été excédentaire pour 38 années sur les 60, soit 63%.

Par ailleurs, on retient aussi que le *Prix du Baril* est lié au *PIB par Habitant* et à la *Balance Commerciale*. En effet, le *Prix du Baril* et le *PIB par Habitant* sont liés positivement, ce qui veut dire l'augmentation du *Prix du Baril* entraîne l'augmentation du *PIB par Habitant* et la variabilité du PIBH est expliquée à 85,77% par la variabilité du PB. Cependant, le *Prix du Baril* et le *Taux de Chômage* ne sont pas significativement liés. En d'autres termes, la variation du *Prix du Baril* n'a pas d'incidence sur le *Taux de Chômage*.

On note pour finir qu'une étude plus approfondie et multidimensionnelle de l'évolution des prix du pétrole au Nigeria de 1960 à 2019 pourrait fournir une compréhension plus complète des défis et des opportunités auxquels le pays est confronté en relation avec son industrie pétrolière.

# ANNEXES

Annexe 1 : JEU DE DONNEES APRES APUREMENT

annees	PB	TXCH	PIBH	CN	TXCA	BC	SB	BP
1960	1.6300	2.430	92.960	0.714	0.72000	defavorable	excedent	equilibre
1961	1.5700	2.410	96.980	0.714	0.19200	defavorable	excedent	equilibre
1962	1.5200	2.440	104.390	0.714	4.10300	defavorable	excedent	equilibre
1963	1.5000	2.470	107.540	0.714	8.57900	defavorable	excedent	equilibre
1964	1.4500	2.500	113.170	0.714	4.95000	defavorable	excedent	equilibre
1965	1.4200	2.450	117.190	0.714	4.88500	defavorable	excedent	equilibre
1966	1.3600	2.370	124.310	0.714	-4.25100	defavorable	excedent	equilibre
1967	1.3300	2.390	99.410	0.714	-7.00905	defavorable	excedent	equilibre
1968	1.3295	2.430	97.200	0.714	-1.24800	defavorable	excedent	equilibre
1969	1.3295	2.510	121.250	0.714	14.29255	defavorable	excedent	equilibre
1970	1.3295	2.500	224.100	0.714	14.29255	defavorable	deficit	desequilibre
1971	1.7000	2.380	160.250	0.713	14.23800	defavorable	deficit	desequilibre
1972	1.8200	2.390	209.230	0.658	3.36400	defavorable	deficit	desequilibre
1973	2.7000	2.380	252.230	0.658	5.39300	favorable	excedent	equilibre
1974	11.0000	2.370	402.850	0.630	11.16100	favorable	excedent	equilibre
1975	10.4300	2.370	166.770	0.616	-5.22800	defavorable	deficit	desequilibre
1976	11.6000	2.370	556.700	0.627	9.04200	defavorable	deficit	desequilibre
1977	12.5000	5.080	536.220	0.645	6.02400	favorable	deficit	desequilibre
1978	12.7900	8.107	527.310	0.635	-5.76400	defavorable	deficit	desequilibre
1979	29.1900	6.720	662.260	0.604	6.75900	favorable	excedent	equilibre
1980	35.5200	4.890	874.400	0.547	4.20500	favorable	excedent	equilibre
1981	34.0000	4.660	2180.200	0.618	-7.00905	defavorable	deficit	desequilibre
1982	32.3800	4.910	1843.910	0.673	-6.80300	defavorable	deficit	desequilibre
1983	29.0400	5.000	1222.630	0.724	-7.00905	favorable	deficit	desequilibre
1984	28.2000	5.900	618.635	0.767	-1.11600	favorable	excedent	equilibre
1985	27.0100	5.000	882.520	0.894	5.91300	favorable	excedent	equilibre
1986	13.5300	4.640	639.010	1.755	0.06100	favorable	excedent	equilibre
1987	17.7300	4.880	598.260	4.016	3.20000	favorable	deficit	desequilibre
1988	14.2400	5.000	549.240	4.537	7.33400	favorable	deficit	desequilibre

annees	PB	TXCH	PIBH	CN	TXCA	BC	SB	BP
1989	17.3100	5.370	474.230	7.365	1.91900	favorable	excedent	equilibre
1990	22.2600	6.010	567.230	8.038	11.77700	favorable	excedent	equilibre
1991	18.6200	3.640	502.910	9.909	0.35800	favorable	excedent	equilibre
1992	18.4400	3.670	477.180	17.298	4.63100	favorable	excedent	equilibre
1993	16.3300	3.740	270.220	22.065	-2.03500	favorable	deficit	desequilibre
1994	15.5300	3.760	321.320	21.996	-1.81500	favorable	deficit	equilibre
1995	16.8600	3.760	408.180	21.895	-0.07300	favorable	deficit	desequilibre
1996	20.2900	3.770	461.520	21.884	4.19600	favorable	excedent	equilibre
1997	18.8600	3.760	479.980	21.886	2.93700	favorable	excedent	equilibre
1998	12.2800	3.760	469.430	21.886	2.58100	defavorable	deficit	desequilibre
1999	17.4400	3.790	497.840	92.338	0.58400	favorable	excedent	equilibre
2000	27.6000	3.780	567.930	101.697	5.01600	favorable	excedent	equilibre
2001	23.1200	3.780	590.380	111.231	5.91800	favorable	excedent	equilibre
2002	24.3600	3.820	741.750	120.578	14.29255	favorable	excedent	equilibre
2003	28.1000	3.820	795.390	129.222	7.34700	favorable	excedent	equilibre
2004	36.0500	3.790	1007.870	132.888	9.25100	favorable	excedent	equilibre
2005	50.5900	3.740	1268.380	131.274	6.43900	favorable	excedent	equilibre
2006	61.0000	3.650	1656.420	128.652	6.05900	favorable	excedent	equilibre
2007	69.0400	3.570	1883.460	128.808	6.59100	favorable	excedent	equilibre
2008	94.1000	3.540	2242.870	118.567	6.76400	favorable	excedent	equilibre
2009	60.8600	3.720	1891.340	148.880	8.03700	favorable	deficit	equilibre
2010	77.3800	3.770	2292.450	150.298	8.00600	favorable	excedent	equilibre
2011	96.8485	3.770	2520.400	153.863	5.30800	favorable	excedent	equilibre
2012	96.8485	3.730	2746.990	157.500	4.23000	favorable	excedent	equilibre
2013	96.8485	3.700	2998.070	157.312	6.67100	favorable	deficit	equilibre
2014	96.2900	4.560	3222.690	158.553	6.31000	favorable	deficit	equilibre
2015	49.4900	4.310	2730.430	192.440	2.65300	defavorable	excedent	desequilibre
2016	40.7600	7.060	2176.000	149.589	-1.61700	defavorable	deficit	equilibre
2017	52.5100	8.107	1968.560	305.790	0.80600	defavorable	deficit	equilibre
2018	69.7800	8.107	2028.180	306.084	1.92300	defavorable	excedent	equilibre
2019	64.0400	8.100	2229.859	306.921	2.20800	defavorable	deficit	desequilibre

*Source : Informations tirées du site [www.banquemondiale.org](http://www.banquemondiale.org), Nos calculs*

*Annexe 2 : Observations contenant des données manquantes*

N°	annees	PB	BC	TXCH	PIBH	CN	SB	TXCA	BP
7	1966	1.36	<NA>	2.37	124.31	0.714	excedent	-4.251	equilibre
16	1975	10.43	defavorable	2.36	NA	0.616	deficit	-5.228	desequilibre
19	1978	12.79	defavorable	10.00	527.31	0.635	deficit	-5.764	<NA>
25	1984	28.20	favorable	5.90	NA	0.767	excedent	-1.116	equilibre
31	1990	22.26	favorable	6.01	567.23	8.038	excedent	11.777	<NA>
35	1994	15.53	<NA>	3.76	321.32	21.996	deficit	-1.815	equilibre
51	2010	77.38	favorable	3.77	2292.45	150.298	<NA>	8.006	equilibre
57	2016	40.76	defavorable	7.06	2176.00	NA	deficit	-1.617	equilibre
59	2018	69.78	<NA>	8.24	2028.18	306.084	excedent	1.923	equilibre

*Source : Informations tirées du site [www.banquemondiale.org](http://www.banquemondiale.org), Nos calculs*

## SOURCE DU CODE R

### #IMPORTATION DU JEU DE DONNEES

```
setwd("D:/BUREAU/INSEDS/COURS/1-ANALYSE STAT DESC I & II/MiniProjet")  
  
petrole = read.table("petrole.csv", header = TRUE, sep = ";", dec = ".", check.names = FALSE,  
stringsAsFactors = TRUE)
```

### #AFFICHAGE DES 5 PREMIERES ET 3 DERNIERES OBSERVATIONS

```
head(petrole,5)  
tail(petrole,3)
```

### #STRUCTURE DU JEU DE DONNEES

```
str(petrole)
```

### #STATISTIQUES SOMMAIRES DES VARIABLES CLES

```
summary(petrole)
```

### #TRAITEMENT DES DOUBLONS

```
petrole_1 = unique(petrole)  
nrow(petrole) - nrow(petrole_1)
```

### # ELABORATION DU DICTIONNAIRE DE DONNEES

```
library("dplyr")  
  
dico_donnees <- tibble::tribble(  
  ~VARIABLE, ~NATURE, ~DESCRIPTION, ~MODALITES,  
  "PB", "Quantitative", "Prix du Baril", "Numerique decimal",  
  "BC", "Qualitative", "Balance Commerciale", "defavorable/favorable",  
  "TXCH", "Quantitative", "Taux de Chomage", "Numerique decimal",  
  "PIBH", "Quantitative", "PIB par Habitant", "Numerique decimal",  
  "CN", "Quantitative", "Cours du Naira", "Numerique decimal",  
  "SB", "Qualitative", "Solde Budgetaire", "excedent / deficit",  
  "TXCA", "Quantitative", "Taux de Croissance Annuel", "Numerique decimal",  
  "BP", "Qualitative", "Balance des paiements", "equilibre / desequilibre")  
  
dico_donnees
```

#### #AFFICHAGE DU DICTIONNAIRE DE DONNEES DANS UN TABLEAU

```
library(rhandsontable)
rhandsontable(dico_donnees, rowHeaders = NULL,
              digits = 3, useTypes = FALSE, search = FALSE,
              width = NULL, height = NULL)
```

#### #VISUALISATION DES VALEURS MANQUANTES

```
library(visdat)
vis_dat(petrole)
vis_miss(petrole)
```

#### #IDENTIFICATION DU NOMBRE D INDIVIDUS AYANT DES VALEURS MANQUANTES

```
petrole[!complete.cases(petrole),]
nrow(petrole[!complete.cases(petrole),])
nrow(petrole[!complete.cases(petrole),]) / nrow(petrole)
```

#### #EXTRACTION DE TOUTES LES VARIABLES QUANTITATIVES

```
quant_vars <- sapply(petrole, is.numeric)
quant <- petrole[, quant_vars]
quant
```

#### #TRAITEMENT DES VALEURS MANQUANTES QUANTITATIVES

```
shapiro.test(petrole$PIBH)
shapiro.test(petrole$CN)
```

```
library(DMwR2)
quant2 <- knnImputation(quant, k = 10, meth = "median")
summary(quant2)
```

#### #EXTRACTION DE TOUTES LES VARIABLES QUALITATIVES

```
qual_vars <- sapply(petrole, is.factor)
qual <- petrole[, qual_vars]
qual
```

### #TRAITEMENT DES VALEURS MANQUANTES QUALITATIVES

```
library(VIM)

qual<- kNN(qual)

qual<- subset(qual, select = BC : BP)

summary(qual)
```

### #CONSTITUTION DU NOUVEAU JEU DE DONNEES SANS VALEURS MANQUANTES

```
petrole_2 = cbind(quant2,qual)

summary(petrole_2)
```

### #VISUALISATION DES VALEURS MANQUANTES APRES TRAITEMENT

```
library(visdat)

vis_dat(petrole_2)

vis_miss(petrole_2)
```

### #VISUALISATION DES VALEURS ABERRANTES ET EXTREMES

```
par(mfrow=c(2,3), mar=c(3,3,3,3))

boxplot(petrole_2$PB, main = "Prix du Baril", col = "red")

boxplot(petrole_2$TXCH, main = "Taux de Chomage", col = "yellow")

boxplot(petrole_2$PIBH, main = "PIB par Habitant", col = "brown")

boxplot(petrole_2$CN, main = "Cours du Naira", col = "green")

boxplot(petrole_2$TXCA, main = "Taux de Croissance Annuel", col = "pink")

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

### #WINZORISATION DES VARIABLES AVEC VALEURS ABERRANTES ET EXTREMES

```
library(DescTools)

petrole_2$PB = Winsorize(petrole_2$PB)

petrole_2$TXCH = Winsorize(petrole_2$TXCH)

petrole_2$TXCA = Winsorize(petrole_2$TXCA)

#VISUALISATION DES VARIABLES APRES TRAITEMENT

par(mfrow=c(2,3), mar=c(3,3,3,3))

boxplot(petrole_2$PB, main = "Prix du Baril", col = "red")

boxplot(petrole_2$TXCH, main = "Taux de Chomage", col = "yellow")

boxplot(petrole_2$PIBH, main = "PIB par Habitant", col = "brown")

boxplot(petrole_2$CN, main = "Cours du Naira", col = "green")

boxplot(petrole_2$TXCA, main = "Taux de Croissance Annuel", col = "pink")

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

## #ANALYSE DESCRIPTIVE UNIVARIEE DE LA VARIABLE PB

### #COURBE D EVOLUTION DU PB

```
library(ggplot2)
data <- data.frame(temps = petrole_2$annees, valeurs = petrole_2$PB)
p <- ggplot(data, aes(x = temps, y = valeurs)) +
  geom_line(color = "red", linetype = 7 ) +
  labs(title = "Courbe d'évolution du Prix du Baril au Nigéria de 1960 à 2019", x = "Année", y = "Prix du
Baril") +
  theme_minimal() +
  theme(plot.title = element_text(size = 14, hjust = 0.5), # Taille du titre
        axis.title = element_text(size = 13), # Taille des titres d'axe
        axis.text = element_text(size = 12), # Taille du texte d'axe
        plot.background = element_rect(fill = "white"), # Fond blanc
        panel.background = element_rect(fill = "white"),
        )
print(p)
```

### #HISTOGRAMME DU PRIX DU BARIL

```
histogramme = hist(petrole_2$PB, main = "Histogramme du PB par Habitant au Nigéria de 1960 à
2019", xlab = "Prix du Baril", ylab = "Effectif (nombre d'années)", ylim = c(0,20), col = "brown")
for (i in 1:length(histogramme$counts)) {
  text(histogramme$mids[i], histogramme$counts[i], labels = histogramme$counts[i], pos = 3)
}
```

### #RESUMES STATISTIQUES DU PB

```
akposso.qt.resume(petrole_2$PB)
```



## #ANALYSE DESCRIPTIVE UNIVARIEE DE LA VARIABLE PIBH

### #COURBE D EVOLUTION DU PIBH

```
library(ggplot2)

data <- data.frame(temps = petrole_2$annees, valeurs = petrole_2$PIBH)

p <- ggplot(data, aes(x = temps, y = valeurs)) +
  geom_line(color = "red", linetype = 7) +
  labs(title = "Courbe d'évolution du PIB par Habitant au Nigéria de 1960 à 2019", x = "Année", y = "PIB
par Habitant") +
  theme_minimal() +
  theme(plot.title = element_text(size = 14, hjust = 0.5), # Taille du titre
        axis.title = element_text(size = 14), # Taille des titres d'axe
        axis.text = element_text(size = 12), # Taille du texte d'axe
        plot.background = element_rect(fill = "white"), # Fond blanc
        panel.background = element_rect(fill = "white"),
  )
print(p)
```

### #HISTOGRAMME DU PIB PAR HABITANT

```
histogramme2 = hist(petrole_2$PIBH, main = "Histogramme du PIBH par Habitant au Nigéria de 1960
à 2019", xlab = "PIB par Habitant", ylab = "Effectif (nombre d'années)", ylim = c(0,30), col = "green")

for (i in 1:length(histogramme2$counts)) {
  text(histogramme2$mids[i], histogramme2$counts[i], labels = histogramme2$counts[i], pos = 3)
}
```

### #RESUMES STATISTIQUES DU PIBH

```
akposso.qt.resume(petrole_2$PIBH)
```

## #ANALYSE DESCRIPTIVE UNIVARIEE DE LA VARIABLE BC

### #TABLEAU STATISTIQUE DE LA BALANCE COMMERCIALE

```
tab_freq=akposso.q1.tableau(petrole_2$BC)
```

```
tab_freq
```

```
write.table(tab_freq,"D:/BUREAU/INSEDS/COURS/1-ANALYSE STAT DESC I &  
II/MiniProjet/tableau_frequence.csv", row.names=TRUE, sep=";",dec=",")
```

### #DIAGRAMME EN BARRES DE LA BALANCE COMMERCIALE

```
class_freq = table(petrole_2$BC)
```

```
graphique = barplot(class_freq,main = "Diagramme en barres de la Balance Commerciale au Nigeria  
de 1960 à 2019", xlab = "Balance Commerciale", ylab = "Effectif (nombre d'années)", ylim = c(0,45),  
col = rgb(0.9,0.9,0.5))
```

```
text(x = graphique, y = class_freq, label = class_freq, pos = 3, cex = 2)
```

### #DIAGRAMME EN SECTEURS DE LA BALANCE COMMERCIALE

```
library(ggplot2)
```

```
pie_data <- data.frame(classes = names(class_freq), frequencies = as.numeric(class_freq))
```

```
pie_chart <- ggplot(pie_data, aes(x = "", y = frequencies, fill = classes)) +
```

```
  geom_bar(stat = "identity", width = 1) +
```

```
  coord_polar("y", start = 0) +
```

```
  theme_void() +
```

```
  geom_text(aes(label = frequencies), position = position_stack(vjust = 0.5), size = 6) +
```

```
  labs(title = "Diagramme en secteurs de la Balance Commerciale au Nigeria de 1960 à 2019") +
```

```
  theme(plot.title = element_text(hjust = 0.5))
```

```
pie_chart + theme(legend.text = element_text(size = 12))
```

## #ANALYSE DESCRIPTIVE UNIVARIEE DE LA VARIABLE SB

### #TABLEAU STATISTIQUE DU SOLDE BUDGETAIRE

```
tab_freq2=akposso.q1.tableau(petrole_2$SB)

tab_freq2

write.table(tab_freq2,"D:/BUREAU/INSSEDS/COURS/1-ANALYSE STAT DESC I &
II/MiniProjet/tableau_frequence4.csv", row.names=TRUE, sep=";",dec=",")
```

### #DIAGRAMME EN BARRES DU SB

```
class_freq = table(petrole_2$SB)

graphique = barplot(class_freq,main = "Diagramme en barres du Solde Budgétaire au Nigéria de 1960
à 2019", xlab = "Solde Budgétaire", ylab = "Effectif (nombre d'années)", ylim = c(0,45), col = "violet")

text(x = graphique, y = class_freq, label = class_freq, pos = 3, cex = 2)
```

### #CAMEMBERT DU SOLDE BUDGETAIRE

```
library(ggplot2)

pie_data <- data.frame(classes = names(class_freq), frequencies = as.numeric(class_freq))

pie_chart <- ggplot(pie_data, aes(x = "", y = frequencies, fill = classes)) +
  geom_bar(stat = "identity", width = 1) +
  coord_polar("y", start = 0) +
  theme_void() +
  geom_text(aes(label = frequencies), position = position_stack(vjust = 0.5), size = 6) +
  labs(title = "Diagramme en secteurs du Solde Budgétaire au Nigéria de 1960 à 2019") +
  theme(plot.title = element_text(hjust = 0.5))

pie_chart + theme(legend.text = element_text(size = 12))
```

## #ANALYSE BIVARIEE DE 2 VARIABLES QUANTITATIVES PB AVEC PIBH

### #NUAGE DE POINTS DE PIBH EN FONCTION DE PB

```
plot(petrole_2$PB,petrole_2$PIBH, main = "Nuage de points", xlab = "Prix du Baril", ylab = "PIB par habitant")  
abline(lm(petrole_2$PIBH~petrole_2$PB))
```

### #RESUMES NUMERIQUES DE PIBH EN FONCTION DE PB

```
cov(petrole_2$PB,petrole_2$PIBH)  
akposso.2qT.liaison(petrole$PB,petrole_2$PIBH)
```

## #ANALYSE BIVARIEE DE 2 VARIABLES QUANTITATIVES PB AVEC TXCA

### #NUAGE DE POINTS DE TXCA EN FONCTION DE PB

```
plot(petrole_2$PB,petrole_2$TXCA, main = "Nuage de points", xlab = "Prix du Baril", ylab = "Taux de Croissance Annuel")  
abline(lm(petrole_2$TXCA~petrole_2$PB))
```

### #RESUSMES NUMERIQUES TCXA EN FONCTION DE PB

```
akposso.2qT.liaison(petrole$PB,petrole_2$TXCA)
```

## #ANALYSE BIVARIEE DE LA VARIABLE PB AVEC LA VARIABLE BC

### #BOITE A MOUSTACHES DU PB PAR NIVEAU DE FACTEUR

```
boxplot(petrole_2$PB~petrole_2$BC, main = "Boîte à moustache")
```

### #RESUMES NUMERIQUES

```
akposso.qLqT.liaison(petrole_2$PB,petrole_2$BC)
```