

Dossier Analyse de données et descriptive M1 ECAP

Analyse des facteurs influençant le taux de pauvreté dans le
monde en 2019

- sur la base d'observations de 69 pays -

Dossier réalisé par:
Isaline HERVE
Juliette GRISON

Année universitaire 2024 - 2025

Sommaire

Introduction	3
1. Présentation des données	4
2. Analyse descriptive	11
3. ACP	21
4. Estimation économétrique et interprétation des résultats	29
5. Modèles Logit / Probit	40
Conclusion	46
Limites	48
Bibliographie	50
Annexe	52

Introduction

Selon la Banque Mondiale, l'extrême pauvreté dans le monde touche environ 700 millions de personnes vivant avec moins de 2.15 dollars par jour en 2022¹. Cela limite leur accès aux ressources essentielles comme l'alimentation saine, la santé, l'éducation et la capacité à avoir un logement convenable. Bien que les efforts pour la réduction de l'extrême pauvreté dans le monde ont été bénéfiques avant la pandémie de COVID-19 – notamment avec l'action des organisations internationales comme les objectifs de développement durable de l'ONU visant à mettre fin à l'extrême pauvreté d'ici 2030, de profondes inégalités subsistent entre les différents pays.

Ces inégalités mettent en évidence la diversité des causes qui influencent la pauvreté et la nécessité d'identifier ces déterminants afin d'orienter au mieux les pouvoirs publics. Ainsi, l'analyse de facteurs économiques, sociaux et institutionnels est essentielle pour mettre en œuvre des politiques adaptées et efficaces. Parmi ces facteurs, des éléments comme le PIB, l'accès à l'électricité, le taux de chômage ou encore le niveau de corruption sont susceptibles de jouer un rôle majeur dans la manière dont la pauvreté se manifeste et persiste.

Dans ce dossier, nous souhaitons analyser les facteurs qui influencent le taux de pauvreté en 2019. La première partie sera consacrée à une présentation détaillée des données qui sera suivie par une analyse descriptive des variables utilisées. Par la suite, nous effectuerons une analyse des composantes principales. Cela nous donnera des informations pertinentes pour poursuivre avec une analyse économétrique des facteurs agissant sur la pauvreté en utilisant la méthode des moindres carrés ordinaires. Enfin, une section séparée sera dédiée à la mise en place d'une régression logistique en réutilisant les variables de notre sujet principal.

¹ World Bank. (2024). *Poverty overview*.

1. Présentation des données

A. Origine et caractéristiques des données

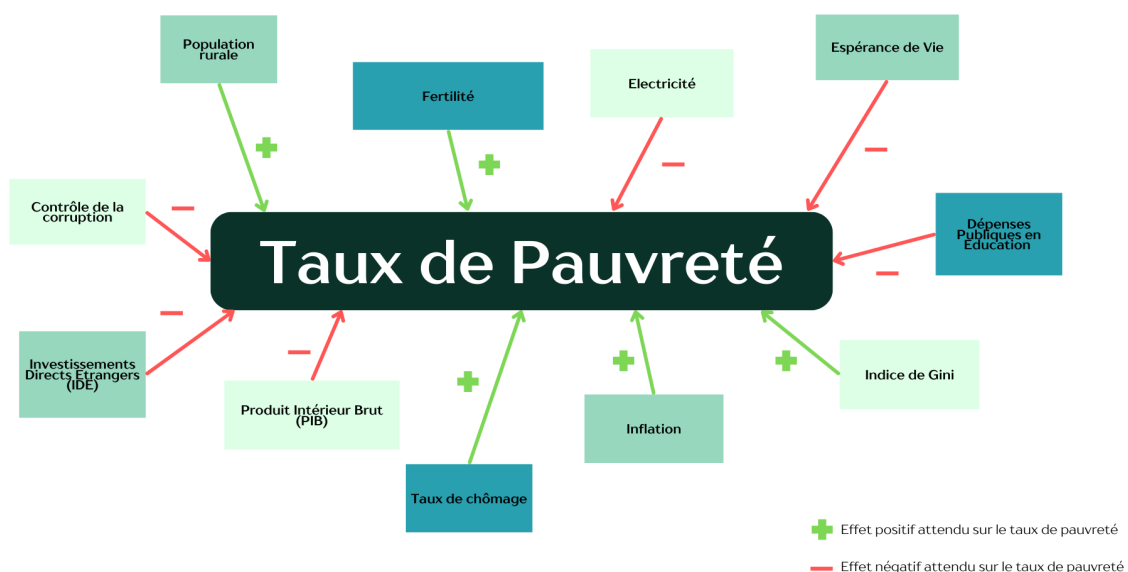
Dans le cadre de notre analyse, nous avons récolté nos données sur le site de la Banque Mondiale.

Premièrement, pour chaque variable, nous avons pris les données relatives à l'année 2019. Le choix de l'année s'est principalement fait en fonction de la disponibilité, 2019 étant l'année la plus récente pour laquelle il y avait le plus de données renseignées. De plus, nous avons pensé que la pandémie du Covid-19 pouvait avoir eu un impact sur de nombreuses variables pour les années suivantes.

Par la suite, en rassemblant les données des variables souhaitées, nous avons remarqué que certaines valeurs étaient manquantes. Nous avons donc convenu d'utiliser la méthode de "Predictive Mean Matching" à l'aide du package {mice} de R afin de conserver le plus d'observations possibles.

Nous nous sommes alors finalement retrouvés avec un échantillon de 69 pays ainsi que 12 variables : 1 variable dépendante quantitative et 11 variables indépendantes, dont 2 qualitatives et 9 quantitatives.

Schéma du modèle de base



B. Présentation des variables par l'analyse documentaire

- **Taux de pauvreté**

Le taux de pauvreté est défini par la Banque Mondiale comme le pourcentage de la population vivant avec moins de 3.65 dollars par jour. Ce ratio est mesuré par rapport aux prix internationaux de 2017 et est donné en pourcentage de la population.

Le taux de pauvreté représente notre variable à expliquer (ou variable dépendante). Nous cherchons à savoir quels éléments ont un impact significatif sur cette variable, ainsi que le sens de cette relation, si elle existe. Les 11 variables explicatives (ou variables indépendantes) qui nous aideront à étudier la pauvreté dans le monde sont présentées en détail ci-dessous.

- **PIB par habitant (en dollars américains courants)**

Le produit national brut (PIB) par habitant représente le PIB divisé par le nombre d'habitants d'un pays donné, calculé au milieu d'une année donnée. Le PIB fait référence à la somme de la valeur ajoutée brute et des taxes sur les produits, moins les subventions qui ne sont pas incluses dans la valeur de ces produits.

Une étude publiée en 2015 montre une relation globalement positive entre la croissance économique et la baisse du niveau de pauvreté. Plus précisément, une relation négative (-0.997) entre le PIB par tête et le taux de pauvreté a été trouvée.²

Nous pouvons en déduire que l'accroissement de la richesse par habitant permet de réduire le taux de pauvreté, et donc, qu'il existe une relation négative entre ces deux facteurs.

- **Indice de Gini**

L'indice de Gini est un indicateur permettant de mesurer la répartition des revenus entre les individus ou les ménages au sein d'un pays donné. Il est particulièrement utile pour évaluer le niveau d'inégalités des revenus d'une économie. Le coefficient de Gini est compris entre 0 (égalité parfaite) et 100 (inégalité absolue).

Selon une étude menée par le groupe de la banque mondiale (World Bank Group) visant à étudier l'impact de la réduction des inégalités sur la pauvreté dans le monde, un scénario dans lequel l'index de Gini diminue de 1% chaque année conduirait à une baisse du taux de

² Henneh, O. F. (2023). *Impact of economic growth on poverty reduction: Evidence from Ghana*. Texila International Journal of Management, 9(2). <https://doi.org/10.21522/TIJMG.2015.09.02.Art007>

pauvreté de 6.3%, soit 89 millions de personnes sortant de la pauvreté extrême.³ En effet, si les richesses sont mieux réparties – c'est-à-dire si l'indice de Gini est faible, la pauvreté relative diminue puisque le revenu médian s'équilibre.

Ainsi, nous supposons une relation positive entre l'indice de Gini et le taux de pauvreté ; un indice de Gini fort indique un niveau élevé de pauvreté.

- **Inflation (en % annuel)**

L'inflation est mesurée par l'indice des prix à la consommation (variation moyenne des prix des produits consommés par les ménages) et représente la variation du prix moyen d'un panier de biens et services sur chaque année. L'indice de Laspeyres a été utilisé (rapport entre la moyenne arithmétique entre les périodes et le coefficient de l'année de base).

Une analyse de données de panel sur 115 pays en développement entre 1981 et 2008 montre que l'inflation est positivement corrélée avec le taux de pauvreté, particulièrement pour les pays à revenu moyen inférieur et les pays à revenu moyen supérieur.⁴ D'autres études comme celle menée au Québec sur l'inflation selon les revenus des familles, indiquent que l'inflation touche particulièrement les familles les plus pauvres. En effet, en ce qui concerne le prix des logements par exemple, le rythme d'inflation est de 31% plus important pour les 20% les plus pauvres que pour les 10% les plus riches sur une période de 20 ans (de 1999 à 2019).⁵

Nous nous attendons donc à ce que l'inflation ait un impact positif sur la pauvreté, quel que soit le niveau de développement du pays donné.

- **Chômage (en % de la population)**

Le chômage représente la part de la population active qui n'a pas d'emploi mais qui en cherche activement et qui est disponible pour travailler. Les spécificités de l'état de chômage d'une personne diffèrent selon le pays.

Selon une étude menée visant à étudier l'influence du chômage sur le niveau de pauvreté en Indonésie, une relation directe et positivement proportionnelle peut être établie entre le chômage et la pauvreté.⁶

³ Lakner, C., Mahler, D. G., Negre, M., & Prydz, E. B. (2020). *How much does reducing inequality matter for global poverty?* World Bank Group.

⁴ Talukdar, S. R. (2012). *The effect of inflation on poverty in developing countries: A panel data analysis* (Master of Arts thesis, Texas Tech University). 1-56.

⁵ Harvey, P.-A., & Nguyen, M. (2020). *L'inégalité face à l'inflation: Une croissance du coût de la vie selon les revenus des familles*. IRIS (Institut de recherche et d'informations socioéconomiques)

⁶ Johan, M., Febrianti, L. A., Wahyuningtyas, I., Salim, F., & Prawira, F. D. (2024). *The influence of unemployment on poverty levels in Indonesia*. ORGANIZE: Journal of Economics, Management and Finance, 3(2), 91–100. <https://doi.org/10.58355/organize.v3i2.65>

Pour notre étude, nous émettons donc l'hypothèse selon laquelle le chômage expose grandement à la pauvreté et que donc, plus le taux de chômage est élevé, plus le niveau de pauvreté est élevé et inversement.

- **Dépenses publiques en éducation (en % du PIB)**

Les dépenses publiques dans l'éducation représentent les dépenses publiques relatives aux établissements d'enseignement publics ou privés, l'administration de l'éducation, ainsi que les subventions à des entités privées tels que les étudiants ou les ménages.

Une étude transnationale publiée en 2019 démontre que les dépenses publiques en éducation réduisent le taux de pauvreté de façon significative. En effet, lors d'une augmentation de 1% de ces dépenses, la pauvreté diminue de 0.227%. Cette relation est négative, quelque soit la sous catégorie des dépenses liées à l'éducation (enseignement primaire, secondaire ou tertiaire).⁷ Cela peut être expliqué par le fait qu'un investissement plus important dans l'éducation donne accès à un nombre d'emplois plus important et mieux rémunérés, ce qui va – à terme, réduire le taux de pauvreté d'un individu ou d'un ménage.

C'est pourquoi nous supposons que les dépenses publiques en éducation auront un impact négatif sur le niveau de pauvreté.

- **Espérance de vie à la naissance (en années)**

L'espérance de vie à la naissance indique la durée de vie moyenne d'un nouveau-né à si les conditions de mortalité l'année de sa naissance restent les mêmes durant toute sa vie.

Une étude réalisée dans le département de Banjarnegara en Indonésie entre 2005 et 2019 montre qu'une augmentation de l'espérance de vie de 1% engendre une diminution du taux de pauvreté de 4.3%.⁸ Cela peut être expliqué par le fait qu'un meilleur accès aux soins de santé, une alimentation suffisante, et des conditions de travail moins pénibles ou dangereuses peuvent réduire les coûts de vie et ainsi diminuer la pauvreté.

Nous nous attendons donc à un effet négatif de l'espérance de vie à la naissance sur le taux de pauvreté.

⁷ Akbar, M., Khan, M., Farooque, H., & Kaleemullah. (2019). *Public spending, education and poverty: A cross-country analysis*. Journal of Multidisciplinary Approaches in Science, 4(1), 12-20. Journal of Multidisciplinary Approaches in Science (JMAS).

⁸ Sudaryati, S., Ahmad, A. A., & Suprpto. (2021). *The effect of average length of schooling, life expectancy, and economic growth on poverty in Banjarnegara Regency 2005-2019*. Eko-Regional: Jurnal Pembangunan Ekonomi Wilayah, 16(1), 29-35. Faculty of Economics and Business, Universitas Jenderal Soedirman.

- **Accès à l'électricité (en % de la population)**

Cet indicateur correspond au pourcentage de la population ayant accès à l'électricité. Pour simplifier notre analyse et la rendre plus pertinente, nous avons transformé cette variable de sorte à ce qu'elle prenne uniquement deux valeurs : "100" et "<100". Ainsi, soit un pays donné dispose d'un accès à l'électricité de 100%, soit il dispose d'un accès à l'électricité inférieur à 100%.

Selon une étude visant à évaluer l'impact de l'accès à l'énergie dans les zones rurales au Cameroun, l'augmentation du prix de l'électricité freine le développement des activités génératrices de revenu. Inversement, l'accès à l'électricité est une condition nécessaire pour ces activités dans les milieux ruraux.⁹ En effet, le fait de ne pas (ou peu) avoir accès à l'électricité peut enfermer une population dans la pauvreté étant donné que l'éducation et le travail ont lieu dans de moins bonnes conditions.

Ainsi, nous supposons que le fait d'avoir un accès à l'électricité de 100% conduira à une diminution du taux de pauvreté.

- **Taux de fertilité (naissances par femme)**

Le taux de fertilité est ici défini comme le nombre moyen d'enfants d'une femme si elle donnait naissance conformément aux taux de fertilité actuels pour chaque tranche d'âge.

Une étude concernant l'inclusion de facteurs démographiques dans l'analyse de la pauvreté dévoile un effet positif de la réduction de la fertilité sur la diminution de la pauvreté sur le long terme. Effectivement, il a été prouvé que la diffusion de nouvelles pratiques de planification familiale ont permis de réduire le taux de pauvreté ainsi que l'écart entre les populations pauvres et celles plus aisées.¹⁰

Nous émettons donc l'hypothèse selon laquelle plus le taux de fertilité augmente, plus le taux de pauvreté augmente à son tour.

- **Population rurale (en % de la population mondiale)**

Le taux de population rurale fait référence aux personnes habitant en milieu rural tel que défini par les bureaux nationaux de statistiques, divisé par la population totale d'un pays donné.

⁹ Kamdem Kamdem, M. (2012). *La contribution de l'énergie à la réduction de la pauvreté en milieu rural au Cameroun*. ICBE-RF Research Report No. 08/12. Investment Climate and Business Environment Research Fund (ICBE-RF), Université de Yaoundé II, Cameroun.

¹⁰ Wietzke, F.-B. (2020). *Poverty, inequality, and fertility: The contribution of demographic change to global poverty reduction*. Population and Development Review, 46(1), 65-99.
<https://doi.org/10.1111/padr.12317>

Selon un rapport de l'IMF (Fonds Monétaire International), 63% de la pauvreté dans le monde provient des milieux ruraux. Dans certains pays – notamment des pays en développement, ce chiffre est encore plus élevé (90% au Bangladesh et jusqu'à 90% dans certains pays d'Afrique subsaharienne).¹¹ Ainsi, il est possible que plus la part de personnes vivant en milieu rural est importante, plus le taux de pauvreté sera élevé.

C'est pourquoi nous nous attendons à une relation positive entre le taux de population rurale et celui de la pauvreté.

- **Contrôle de la corruption**

L'indicateur de contrôle de la corruption reflète la mesure dans laquelle le gouvernement d'un pays donné exerce son pouvoir à des fins privées ainsi que la "capture" du gouvernement par des élites ou des groupes privés. C'est un indicateur global de perception de corruption, utilisant une échelle normale et standardisée allant de -2.5 à 2.5, -2.5 étant très corrompu et 2.5 étant très peu corrompu.

Un article publié en 2020 à propos de la corruption en Afrique affirme que la pauvreté est perpétuée par une corruption croissante, notamment à travers ses effets sur la réduction de la crédibilité des institutions publiques.¹² Cela peut s'expliquer par le fait qu'un gouvernement plutôt corrompu risque de ralentir la croissance économique d'un pays donné, et de détourner les fonds des services publics essentiels, limitant ainsi la capacité des plus pauvres à emprunter ou augmenter leurs revenus.

Nous supposons donc qu'un niveau de corruption élevé (ou un indicateur de contrôle de corruption faible) conduit à un taux de pauvreté élevé.

- **Investissements Directs Étrangers (en % du PIB)**

Les investissements directs étrangers (IDE) correspondent aux entrées nettes d'investissement effectués dans le but d'acquérir un intérêt durable, dans une économie différente de celle de l'investisseur. Ici, cet indicateur représente les flux entrants nets des investisseurs étrangers dans l'économie déclarante. Nous avons fait le choix de transformer cette variable initialement numérique en une variable binaire prenant 2 valeurs : "Négatif" et "Positif", faisant chacune référence au signe de l'IDE du pays donné.

¹¹ Khan, M. H. (2002). *La pauvreté rurale dans les pays en développement: Orientation pour l'action publique*. Economic Issues Series, No. 26. International Monetary Fund.

¹² Eshun, J., & Baah, J. A. (2019). *What does recent survey data say about the effect of corruption on poverty in Africa?* Journal of Economics, Management and Trade, 25(6), 1-13.

Selon un article publié en 2021 par l'IJAFAME, la majorité des études empiriques sur le sujet montrent qu'une augmentation des IDE permet de réduire la pauvreté dans les économies hôtes, soulignant le rôle positif que l'implémentation de firmes étrangères a sur la diminution et la limitation du taux de pauvreté dans les pays d'accueil.¹³

C'est dans ce contexte que nous pensons qu'il existe une relation négative entre l'IDE et le taux de pauvreté.

C. Résumé des variables du modèle

Tableau récapitulatif des variables explicatives

Variable	Correspondance sur R	Type	Effet attendu
PIB par habitant	<i>pib</i>	Quantitative	Négatif (-)
Chômage	<i>chomage</i>	Quantitative	Positif (+)
Inflation	<i>inflation</i>	Quantitative	Positif (+)
Investissements Directs Étrangers	<i>ide</i>	Qualitative binaire : - "Négatif" - "Positif"	Négatif (-)
Indice de Gini	<i>indice_gini</i>	Quantitative	Positif (+)
Espérance de vie à la naissance	<i>esp_vie</i>	Quantitative	Négatif (-)
Accès à l'électricité	<i>electricite</i>	Qualitative binaire : - "< 100" : accès à l'électricité inférieur à 100% - "100" : accès à l'électricité de 100%	Négatif (-)
Taux de fertilité	<i>fertilite</i>	Quantitative	Positif (+)
Population rurale	<i>pop_rurale</i>	Quantitative	Positif (+)
Contrôle de la corruption	<i>corruption</i>	Quantitative	Négatif (-)
Dépenses publiques en éducation	<i>dep_educ</i>	Quantitative	Négatif (-)

¹³ Moujahid, M., & Khariss, M. (2021). *Investissement direct étranger et réduction de la pauvreté : Essai d'élaboration d'une revue de la littérature*. International Journal of Accounting, Finance, Auditing, Management and Economics, 2(4), 356-373. <https://doi.org/10.5281/zenodo.5138765>

2. Analyse descriptive

La prochaine section de ce dossier est consacrée à l'analyse descriptive de nos données. Elle sera divisée en deux parties : l'analyse univariée puis, l'analyse bivariée.

A. Analyse univariée

1. Variables qualitatives

Dans le cadre de ce projet, nous avons deux variables qualitatives, la variable électricité et la variable IDE, toutes deux binaires. La variable électricité prend deux valeurs : le fait d'avoir un taux d'accès à l'électricité ou non. La variable IDE prend deux valeurs : le fait d'avoir un IDE positif ou négatif. Représentons maintenant visuellement ces modalités.

Diagrammes circulaires des variables qualitatives

Diagramme circulaire de la variable electricite

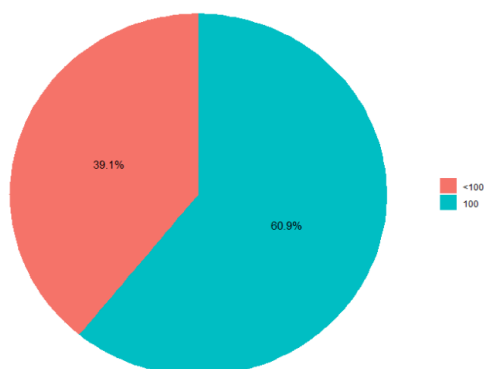
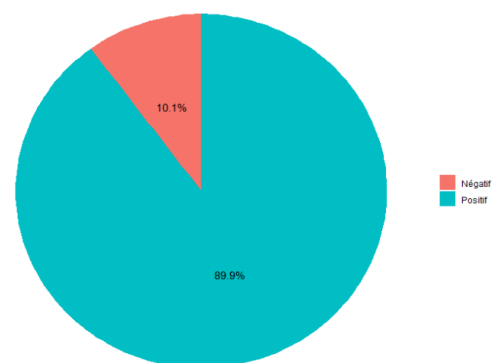


Diagramme circulaire de la variable ide



Comme nous pouvons le voir, les modalités de la variable électricité semblent plutôt bien réparties, avec environ 60% des pays ayant un taux d'accès à l'électricité de 100% et 40% en ayant un inférieur à 100%.

Concernant les investissements directs étrangers cependant, beaucoup plus de pays ont une valeur positive (environ 90%) que négative (environ 10%).

2. Variables quantitatives

a. Indicateurs de tendance centrale et de dispersion

En premier lieu, nous avons représenté les mesures de tendance centrale (moyenne, médiane) et de dispersion (minimum, maximum, écart-type) de nos 10 variables quantitatives.

Tableau de statistiques descriptives des variables quantitatives

Variable	Moyenne	Médiane	Minimum	Maximum	Ecart-type
pauvrete	9.77	1.3	0.1	89.1	19.4
pib	20898	10144	512	112726	22856
chomage	6.60	5.55	0.716	17.4	3.63
inflation	6.82	2.10	-3.55	255	30.8
indice_gini	35.5	34.2	23.2	53.5	6.94
dep_educ	4.52	4.25	0.964	8.09	1.45
esp_vie	76.4	76.8	61.2	83.8	5.62
fertilite	2.00	1.7	0.918	4.81	0.859
pop_rurale	33.8	31.9	1.96	82.8	18.9
corruption	0.169	0.00690	-1.56	2.12	1.06

Ce tableau nous montre des minimums négatifs pour les variables inflation et corruption, indiquant qu'une ou plusieurs observations ont une valeur négative, ce qui n'est pas anormal étant donné les unités de ces variables (l'inflation est en pourcentage et les valeurs de corruption peuvent se situer entre -2.5 et 2.5). Il est tout de même important de noter le maximum de 255 pour la variable inflation, qui n'a pas de sens dans ce contexte. Il est possible que ce soit une erreur de saisie ou une information mal renseignée. Dans tous les cas, elle peut constituer une valeur atypique et peut fausser nos statistiques en augmentant la valeur de l'écart-type.

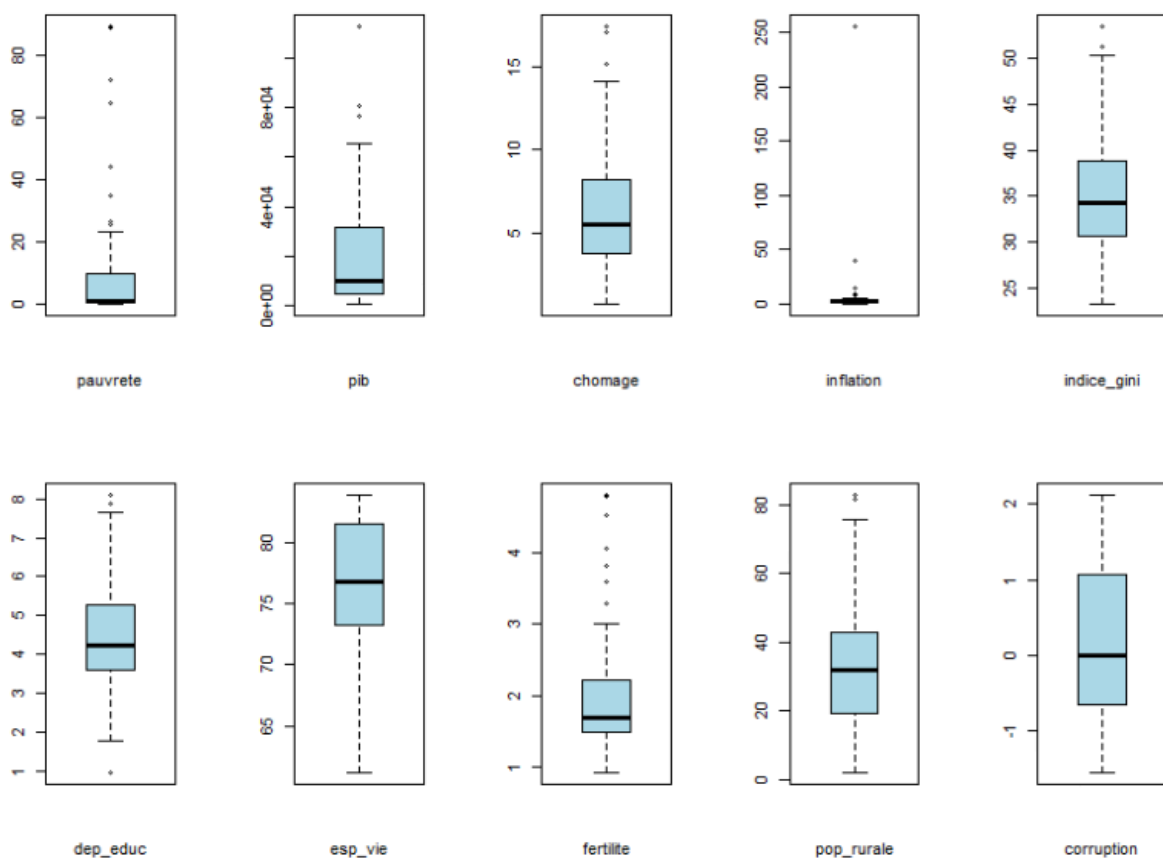
En outre, nous pouvons voir que la moyenne et la médiane de la majorité des variables sont relativement proches, ce qui est une bonne indication. En effet, cela signifie que la distribution des données est plutôt symétrique et n'est pas beaucoup influencée par les valeurs extrêmes. Ce n'est cependant pas le cas pour les variables de la pauvreté et du PIB,

avec un écart entre ces deux indicateurs assez important, ce qui indique potentiellement l'existence d'une ou plusieurs valeurs atypiques.

Ces quelques observations et interrogations nous mènent à nous pencher plus en détail sur la présence ou non de valeurs atypiques. Celles-ci constitueraient un problème lors de notre analyse, en raison du biais qu'elles pourraient créer, menant à une interprétation faussée des résultats trouvés. Nous avons d'abord choisi de les visualiser à l'aide de boîtes à moustache.

b. Visualisation et traitement de valeurs atypiques

Boîtes à moustache des variables quantitatives



Ici, quelques valeurs atypiques potentielles sont aperçues, notamment au sein des variables pauvreté, PIB, chômage, inflation, indice de Gini, dépenses en éducation, fertilité et population rurale.

Pour détecter leur réelle atypicité, nous avons effectué un test de Rosner sur chacune d'entre elles. A la suite de ce test, nous avons trouvé qu'il existe 7 valeurs atypiques dans pauvreté, 1 dans PIB et 5 dans inflation, comme nous l'avons suspecté précédemment. Il y en a également 6 dans fertilité.

Finalement, ces valeurs atypiques concernent 12 observations que nous avons choisi de supprimer pour la suite de notre analyse. Nous passons donc de 69 observations à 57 observations.

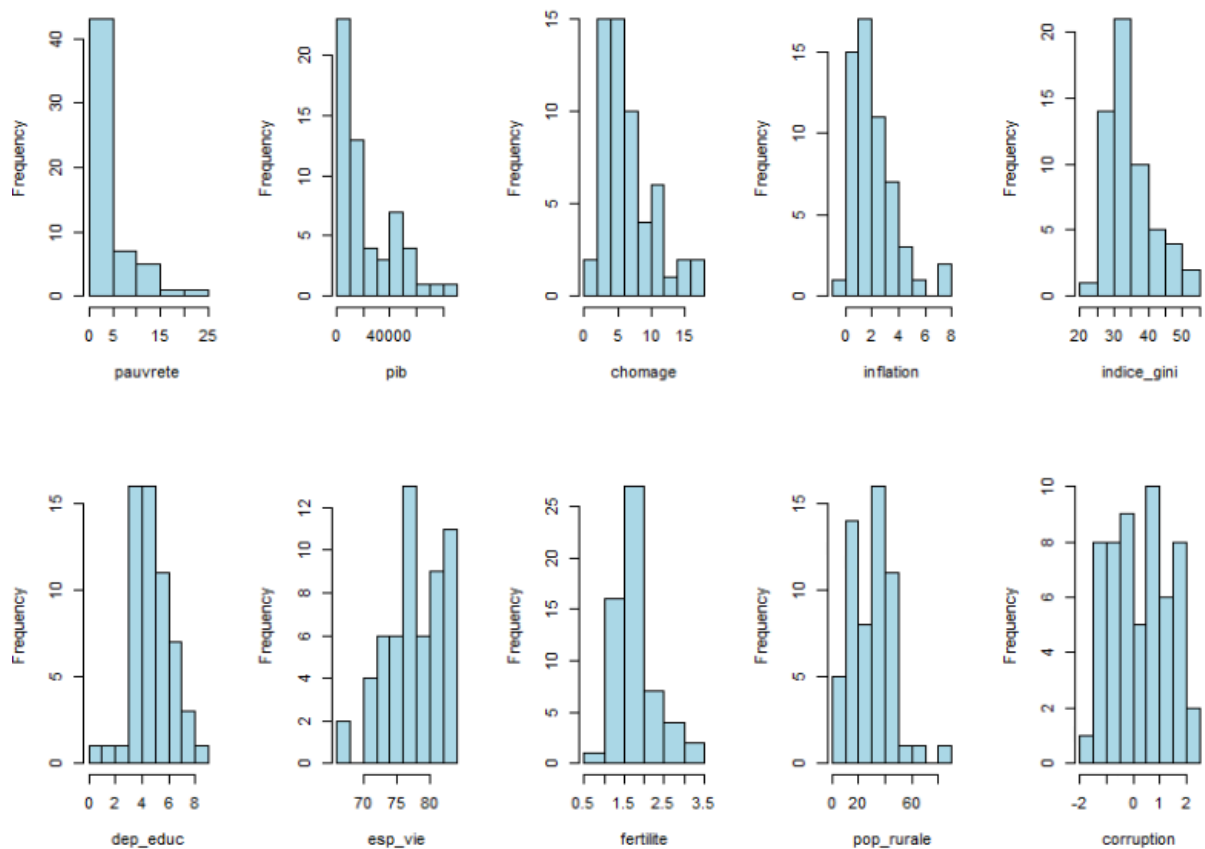
c. Caractéristiques des distributions

Symétrie et aplatissement des variables quantitatives

Skewness et Kurtosis des variables quantitatives

Variable	Skewness	Kurtosis
pauvrete	1.9654313	3.9175448
pib	1.1278192	0.4190371
chomage	1.2794762	1.1695807
inflation	1.5021374	2.9923405
indice_gini	1.0155933	0.7326944
dep_educ	0.2608465	0.4966326
esp_vie	-0.4418703	-0.5472996
fertilite	1.3972933	1.9348153
pop_rurale	0.6390786	1.2008704
corruption	0.0620253	-1.1320146

Histogrammes des variables quantitatives



Ce tableau ainsi que ces histogrammes nous permettent de visualiser la répartition des observations sur les variables quantitatives.

En premier lieu, la variable pauvreté est fortement asymétrique ($\text{skewness} > 0$) avec une queue étalée vers la droite. Cette queue est importante ($\text{kurtosis} > 3$) et non aplatie, indiquant la présence de valeurs extrêmes (bien que ces histogrammes soient créés après le traitement des valeurs atypiques), en plus d'avoir une distribution ne suivant pas une loi normale.

En outre, les variables PIB, chômage, indice de Gini et fertilité présentent les mêmes propriétés que la variable pauvreté ($\text{skewness} > 0$), seulement leur distribution est aplatie ($\text{kurtosis} < 3$). Là aussi se pose le problème d'asymétrie, avec une distribution qui ne suis pas une loi normale.

Enfin, de façon plus subtile, nous observons que le reste des variables quantitatives semblent elles aussi ne pas avoir une distribution normale, avec des coefficients de symétrie et d'aplatissement qui diffèrent respectivement de 0 et de 3.

Pour résoudre ce problème d'asymétrie et de valeurs extrêmes, une solution envisageable serait une transformation de ces variables en forme logarithme, afin de réduire l'échelle de leur distribution, et que celle-ci s'apparente davantage à une loi normale. Cette potentielle transformation n'inclut évidemment pas les variables indice de Gini et corruption étant donné que certaines de leurs valeurs sont négatives. Cependant, nous préservons la forme actuelle de toutes nos variables pour le moment, le choix d'effectuer cette transformation ou non étant l'objet d'une section de notre analyse économétrique ultérieure.

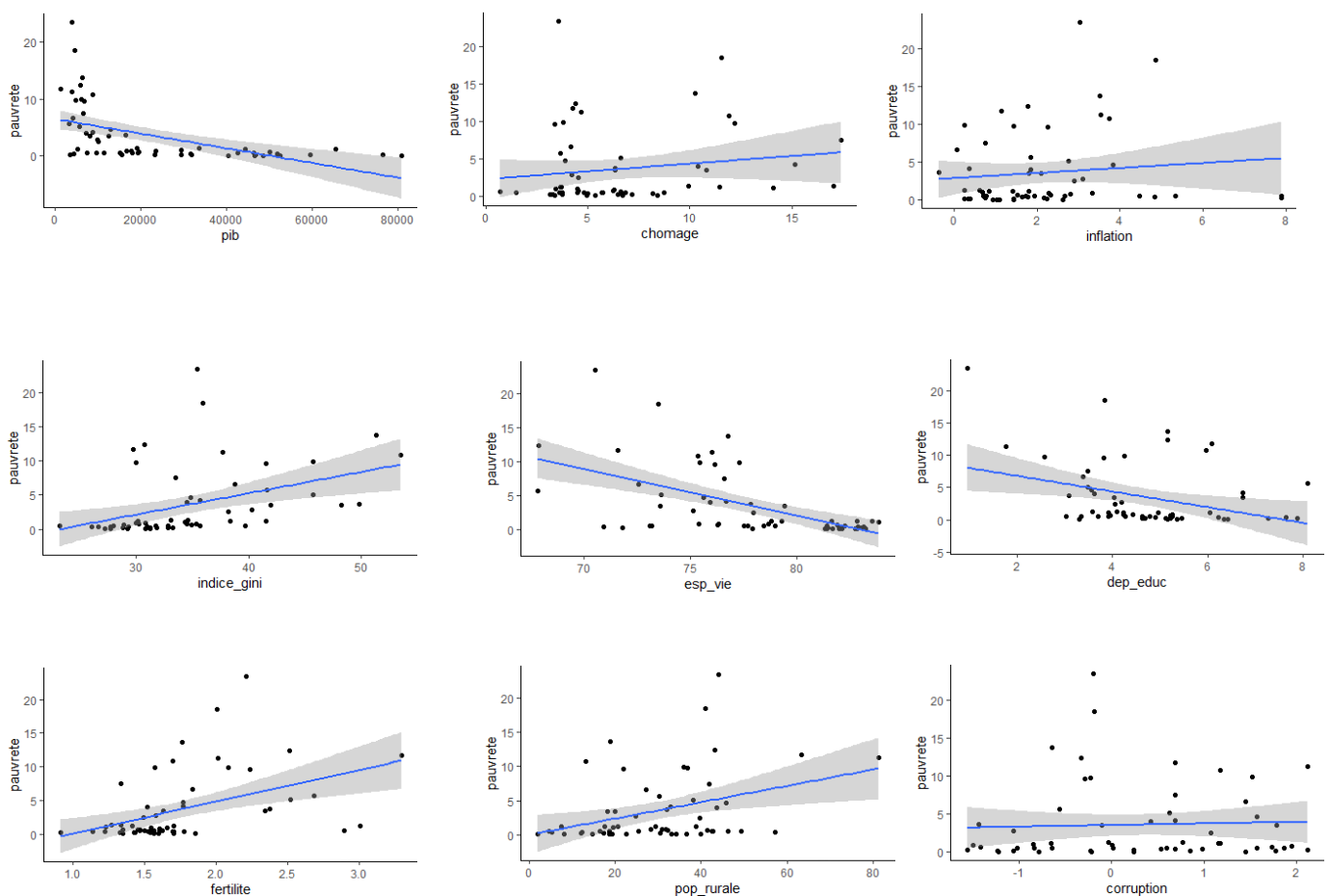
B. Analyse bivariée

1. Quanti-quanti

a. Nuages de points

Nous avons représenté le lien entre nos variables quantitatives et la variable pauvreté avec des nuages de points et des régressions linéaires avec un intervalle de confiance autour de la ligne.

Nuages de points entre variables quantitatives

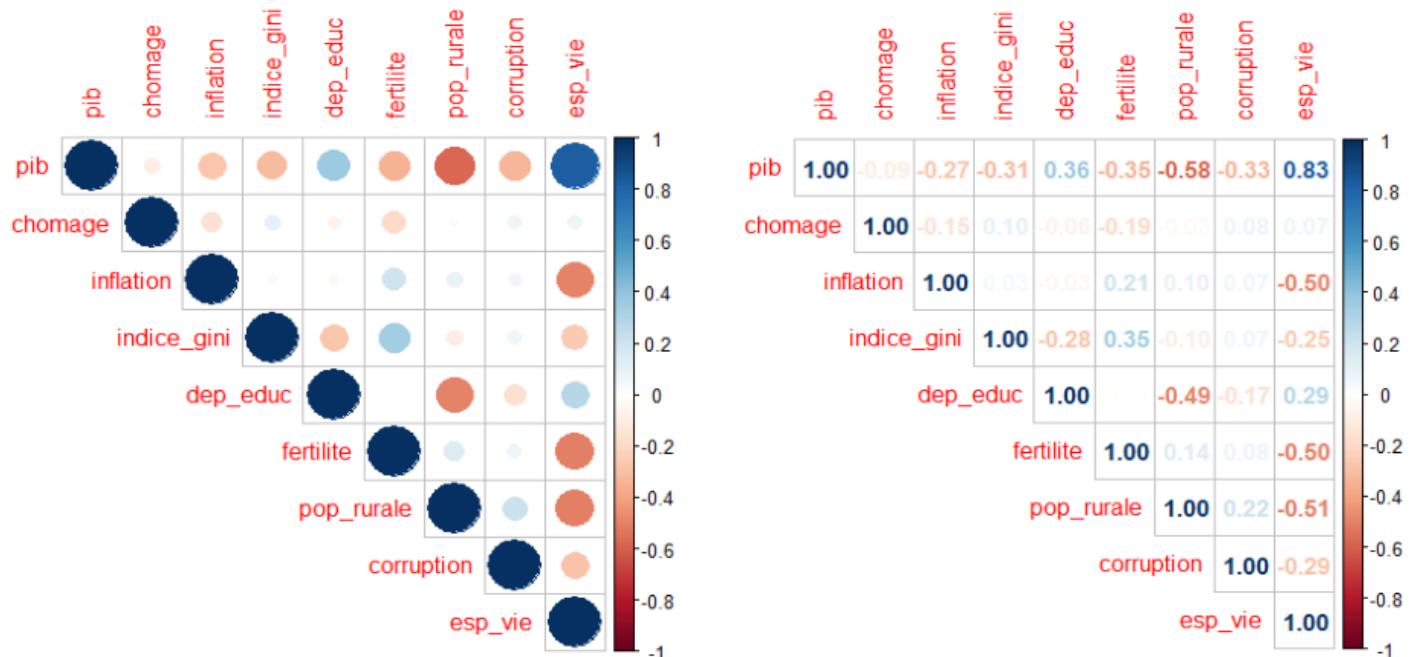


Comme nous pouvons le voir, il y a une relation positive entre les variables “indice_gini”, “fertilité”, “pop_rurale” et la variable “pauvreté”. Une relation légèrement positive entre “chomage”, “inflation” et la variable “pauvreté” peut également être observée. Les variables “pib”, “dep_educ”, et “esp_vie” sont négativement corrélés avec “pauvreté”. Enfin, il n’y pas de corrélation particulière entre “corruption” et pauvreté”. Ces relations – corruption exclue, sont attendues de notre part et de la littérature déjà existante sur le sujet.

b. Corrélations

Nous allons maintenant pouvoir visualiser le signe et la puissance des corrélations entre les variables explicatives quantitatives.

Graphiques des corrélations de Spearman



Ici, nous constatons une corrélation négative modérément forte entre “pop_rurale” et “pib” (-0.58), et entre “pop_rurale” et “dep_educ” (-0.49). La variable “esp_vie” est également modérément corrélée négativement avec : “inflation” (-0.5), fertilité (-0.5) et “pop_rurale”

(0.51). Ces corrélations sont d'ampleur raisonnable, cependant, nous remarquons une très forte corrélation entre “esp_vie” et “pib” (0.83). L'espérance de vie étant déjà corrélée avec d'autres variables, il est important de prendre en compte le fait qu'elle puisse potentiellement introduire un biais dans notre analyse en raison du risque de multicolinéarité.

En outre, le test de Pearson (test de nullité du coefficient de corrélation), permet de vérifier si cette corrélation est statistiquement significative. Après l'avoir effectué, nous trouvons une p-value de 3.179e-11 (p-value < 0.05) ce qui montre qu'il y a une forte relation linéaire positive et significative entre le PIB et l'espérance de vie, au seuil de risque de 5%.

2. Quali-quali

a. Tableau de contingence

Tableau de contingence des variables qualitatives

Electricité / IDE	Négatif	Positif	Total
<100	1	18	19
100	5	38	43
Total	6	56	62

Ce tableau de contingence permet de voir la répartition des observations des deux variables qualitatives : électricité et IDE. Pour savoir si ces deux variables sont indépendantes, nous pouvons vérifier si un pays avec un taux d'accès à l'électricité a autant de chances d'avoir un IDE positif que négatif.

La probabilité de “100” est de 69% (43/62) et la probabilité de “Positif” est de 90% (56/62). Pour que ces variables soient indépendantes, il faut que la probabilité d'être un pays avec “100” et “Positif” soit de **62.1%** (69% × 90%). Nous trouvons que cette probabilité est de **61.3%** (38/62). Ces probabilités étant très similaires, nous pouvons supposer une indépendance entre ces variables.

b. Test de Khi-deux

Pour renforcer notre hypothèse d'indépendance, nous avons effectué un test de Khi-deux. Nous avons trouvé une p-value de 0.6471 (p-value > 0.05) ne nous permettant pas de rejeter

l'hypothèse nulle ; il n'y a donc pas de preuves statistiquement significatives d'une association entre les deux variables. Cela confirme notre hypothèse précédente.

c. V de Cramér

Il est aussi possible d'évaluer cette indépendance grâce au V de Cramér. Nous trouvons un résultat de 0.06 qui est assez proche de 0 et montre encore une fois qu'il n'existe pas de lien (ou un lien extrêmement faible) entre le taux d'accès à l'électricité et l'IDE.

3. Quanti-quali

a. Tests de comparaison des moyennes

Nous cherchons maintenant à tester s'il y a une différence de moyenne significative liée à une réelle différence entre les groupes. Nous effectuons ces tests pour les deux groupes de la variable "electricite", et pour les deux groupes de la variable "ide", par rapport à la variable "pauvrete".

Pour l'électricité, nous trouvons une p-value de 0.0007 (p-value < 0.05), et pour l'IDE, une p-value de 0.0385 (p-value < 0.05). Nous rejetons donc l'hypothèse selon laquelle les moyennes de "pauvrete" entre les deux groupes sont égales au seuil de risque de 5% : il y a donc une différence significative entre les moyennes de pauvreté pour les deux groupes de "electricite" et pour les deux groupes de "ide".

Nous avons uniquement comparé les moyennes de la variable à expliquer "pauvreté" car nous ne sommes pas dans un cadre expérimental avec des groupes de traitement et de contrôle. En effet, nous cherchons juste à modéliser la relation entre les variables explicatives et la variable dépendante (pauvreté). Un test de comparaison de moyennes pour les variables explicatives quantitatives n'est donc pas pertinent ici.

b. Statistique éta carré

Tableau récapitulatif des valeurs de la statistique éta carré

	electricite	ide
pauvrete	0.28	0.03
pib	0.28	0.20

chomage	0.01	0.03
inflation	0.01	0.01
indice_gini	0.45	0.03
dep_educ	0.04	0.09
esp_vie	0.30	0.02
fertilite	0.20	0.001
pop_rurale	0.03	0.03
corruption	0.002	0.07

Nous allons conclure notre analyse descriptive en calculant la statistique éta carré, qui permet de mesurer la force de l'association entre nos variables qualitatives et quantitatives. Selon les balises de Cohen (1988), 28% des variations de la pauvreté sont expliquées par l'accès à l'électricité, ce qui représente un effet fort (> 0.14), tandis que l'effet de l'IDE sur la variation de la pauvreté est de 3%, ce qui représente une association faible (comprise entre 0.01 et 0.06).

Par ailleurs, l'électricité a un effet fort sur l'indice de Gini avec 45% des variations expliquées par l'électricité soit presque la moitié, ainsi qu'un effet fort sur le PIB (28%), l'espérance de vie (30%) et la fertilité (20%). Enfin, 20% des variations du PIB sont expliquées par les variations de l'IDE et est la seule variable étant fortement affectée par l'IDE.

Nous allons maintenant poursuivre avec l'analyse des composantes principales (ACP).

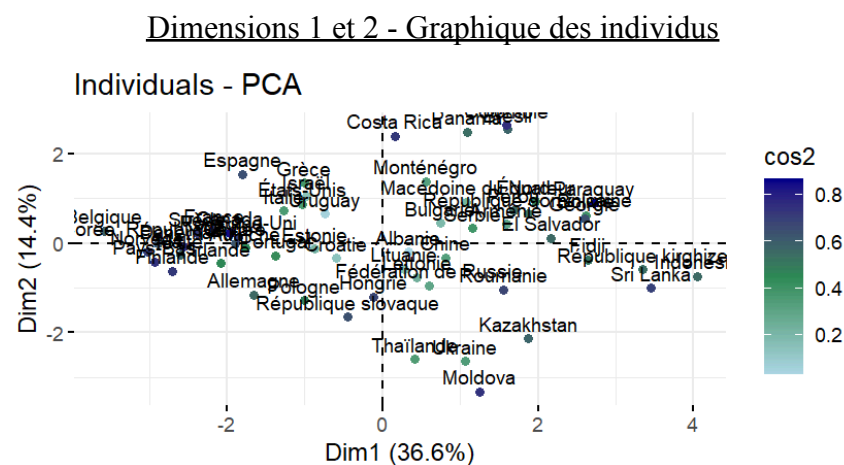
3. ACP

Dans le but d'explorer nos données par rapport aux variables quantitatives de notre modèle, nous avons réalisé une Analyse en Composantes Principales.

A. Analyse simple

Pour commencer, nous avons alors effectué une ACP en sélectionnant seulement les variables quantitatives de notre modèle. Nous avons ensuite comparé les résultats avec une ACP où ces variables étaient en format logarithmiques pour celles que nous avons supposé logarithmique d'après l'analyse descriptive. Les résultats variant peu, nous avons quand même décidé d'interpréter les résultats de l'ACP faite avec les variables logarithmiques, afin d'avoir une analyse plus complète.

Dans un premier temps, nous avons observé le graphique des individus, montrant les observations de nos données dans les dimensions 1 et 2.



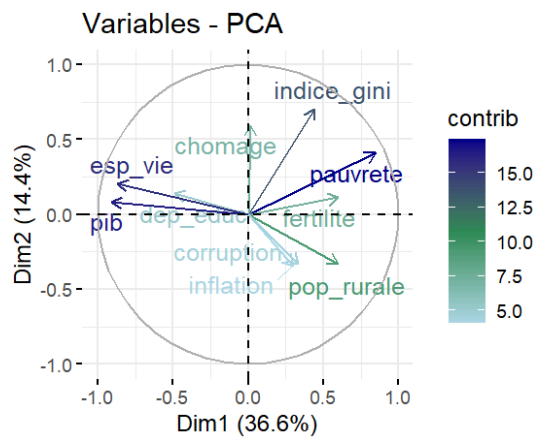
Nous pouvons constater que les observations sont plutôt bien réparties autour du centre ; l'analyse parvient alors à bien les discriminer. Certaines observations contribuent néanmoins davantage à la variable expliquée, comme par exemple l'Indonésie, située loin du centre et proche du cercle de corrélation.

De plus, nous pouvons observer certains regroupements dans les coins supérieur droit, supérieur gauche et inférieur gauche, indiquant potentiellement des groupes partageant des caractéristiques similaires par rapport aux variables de notre jeu de données.

Les 2 premiers axes représentés sur ce graphique contribuent à presque 51% de la variance expliquée cumulée, ce qui est suffisant pour que nous puissions interpréter l'ACP.

Après cela, le graphique des variables nous indique l'influence de chaque variable dans ces mêmes dimensions principales.

Dimensions 1 et 2 - Graphique des variables

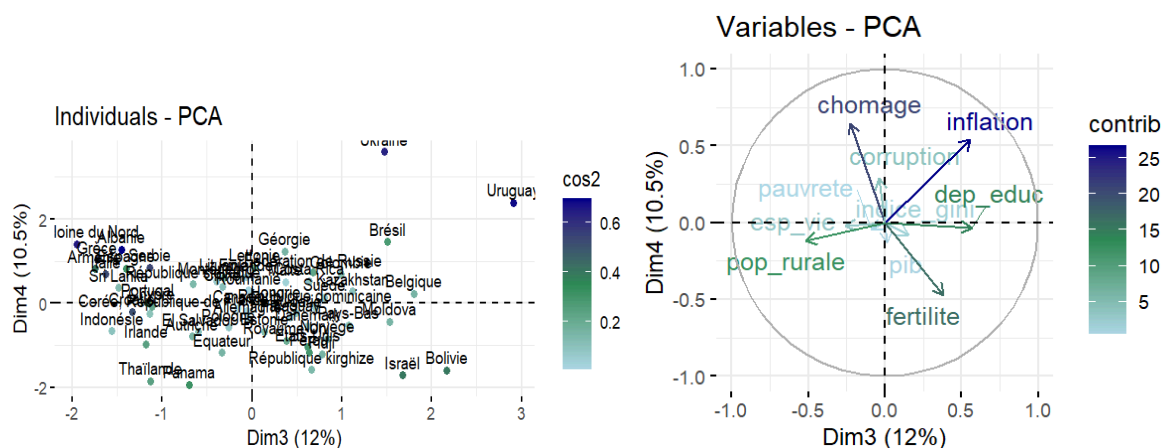


Nous pouvons voir que l'espérance de vie et le PIB sont fortement corrélées négativement avec l'axe 1 ; ces variables étant fortement corrélées entre elles. Ensuite, la variable de la pauvreté influencera positivement la dimension 1, ainsi que la dimension 2.

En ce qui concerne la dimension 2, nous pouvons voir que les principales influences seront les variables de l'indice de Gini et du chômage, corrélées positivement à cet axe.

Nous avons également observé ces graphiques pour les dimensions 3 et 4.

Dimensions 3 et 4 - Graphique des individus et graphique des variables



D'abord, pour le graphique des individus, nous pouvons voir un regroupement des observations en haut à gauche, ainsi qu'un autre regroupement sur la partie droite près du centre.

Le graphique des variables nous indique que les variables participant le plus à la conception de l'axe 3 seront l'inflation, les dépenses en éducation, et la fertilité, avec une influence positive, puis la population rurale, avec une corrélation négative. Pour la dimension 4, nous retrouverons la variable du chômage avec une forte influence positive, ainsi que l'inflation et la fertilité, avec une influence négative.

Ensuite, en examinant les valeurs propres ainsi que le pourcentage de la variance expliquée cumulée pour chaque dimension, nous avons conclu que nous retiendrons dans notre analyse les axes allant de 1 à 4. Effectivement, ces dimensions sont celles ayant des valeurs propres supérieures à 1 et en se basant sur le critère de Kaiser, ce sont donc celles que nous devons garder. La variance explicative cumulée pour ces axes sera de 73,45%.

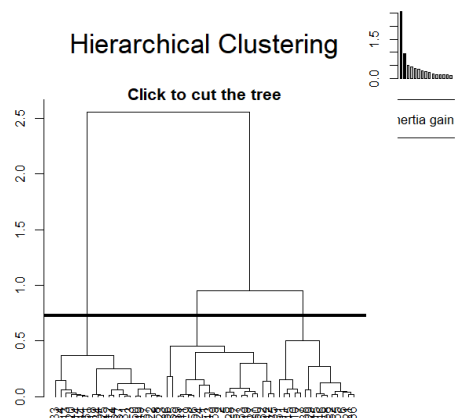
B. Analyse des clusters

A la suite de cela, nous avons mené une analyse en fonction des regroupements de nos observations, dans l'objectif de déduire des classes de pays ayant des caractéristiques similaires.

1. Dimensions 1 et 2

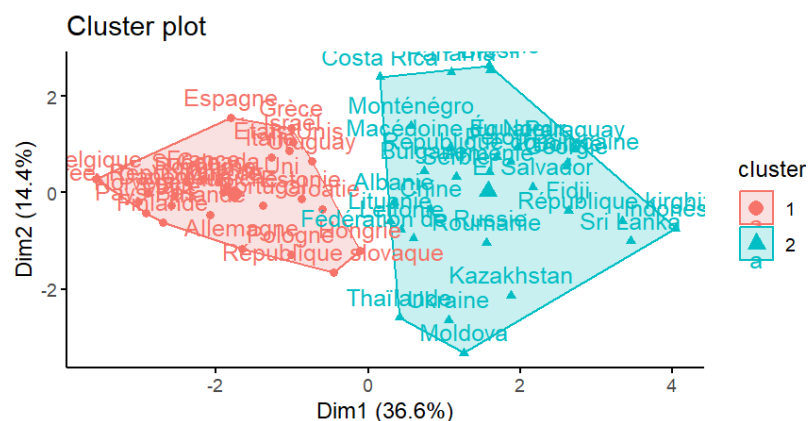
Dans un premier temps, nous avons effectué une classification hiérarchique à partir d'un arbre de classification, d'après le principe d'inertie inter-classe utilisé dans la fonction HCPC du package {FactoMineR}.

Découpage de l'arbre de classification hiérarchique



Avec une ligne de découpage située à 0,75, l'inertie, donc la différence, intra-classe sera faible tandis que l'inertie inter-classe sera forte, indiquant une certaine hétérogénéité entre les groupes qui seront formés. Néanmoins, il y aura tout de même une certaine distinction entre ces groupes, regroupant chacun des observations avec des caractéristiques similaires différentes par rapport à celles des autres groupes.

Dimensions 1 et 2 - Graphique des clusters



Apparaissent alors 2 clusters dont nous allons d'abord analyser les caractéristiques respectives.

Tableau de résumé de la structure des clusters

Variables	Cluster 1 (valeur moyenne)	Cluster 2 (valeur moyenne)	Test ANOVA (différence significative)
pauvrete	-0,99	1,28	Oui

pib	10,47	8,85	Oui
chomage	1,69	1,75	Non
inflation	1,65	2,54	Oui
indice_gini	31,43	37,52	Oui
dep_educ	1,62	1,38	Oui
esp_vie	4,4	4,31	Oui
fertilite	0,41	0,62	Oui
pop_rurale	2,88	3,53	Oui
corruption	-0,2	0,66	Oui

D'après les tests ANOVA à un facteur effectués sur nos variables quantitatives, nous trouvons que 9 variables montrent bien des différences significatives entre les clusters, avec des p-values inférieures à 0,05. Néanmoins, la variable du chômage, n'en montre pas, car la p-value du test est supérieure à 0,05. Il faut également noter que le test est seulement légèrement significatif pour la variable de l'inflation, avec une p-value de 0,046, montrant alors que l'inflation ne varie pas beaucoup entre les groupes.

Ensuite, nous avons également observé l'influence des variables quantitatives sur la classification, d'après la statistique d'Eta². Mesurant la proportion de variance d'une variable expliquée par les clusters, cela nous indique le niveau d'implication de la variable dans la distinction des clusters.

Ainsi, la variable du PIB est celle agissant le plus sur la distinction des clusters, avec un Eta² de 0,7088 et une p-value inférieur à 0,0001.

Après cela, nous retrouvons dans l'ordre la variable de l'espérance de vie, le niveau de pauvreté, la population rurale et l'indice de gini.

Logiquement, les variables ayant contribué le plus à la séparation des groupes sont celles étant le plus corrélées avec les dimensions 1 et 2.

Ensuite, nous retrouvons les autres variables, qui restent significatives (avec des p-values inférieures à 0,05) mais avec une influence davantage modérée.

Ainsi, cela nous permet de diviser nos observations en 2 groupes distinct :

- Le premier groupe, correspond à des pays avec un niveau de pauvreté faible, un PIB plutôt élevé, des inégalités économiques modérés représentées par l'indice de Gini, de fortes dépenses en éducation, une espérance de vie élevée, un taux de fertilité faible, une proportion de la population rurale faible ainsi qu'un indice de corruption très faible.
- Le 2ème groupe correspond au contraire. Nous retrouvons donc un niveau de pauvreté plus élevé, un PIB plus faible, davantage d'inégalités économique avec un indice de Gini plus élevé, moins de dépenses en éducation, une espérance de vie plus faible, un taux de fertilité plus élevé, un nombre plus important de la population étant rurale et un indice de corruption perçu plus élevé.

Le premier cluster représente donc les pays les plus développés économiquement de notre échantillon. Nous retrouvons par exemple la Pologne, le Portugal, la Belgique, l'Irlande, l'Allemagne ou encore l'Espagne.

Le deuxième cluster représente alors les pays en développement, tels que la Bulgarie, la Serbie, la République kirgizhe, le Sri Lanka, la Macédoine du Nord et El Salvador.

2. Dimensions 3 et 4

Étant donné que d'après le critère de Kaiser, nous retenons les dimensions allant de 1 à 4, nous avons également souhaité observer les potentiels clusters des dimensions 3 et 4.

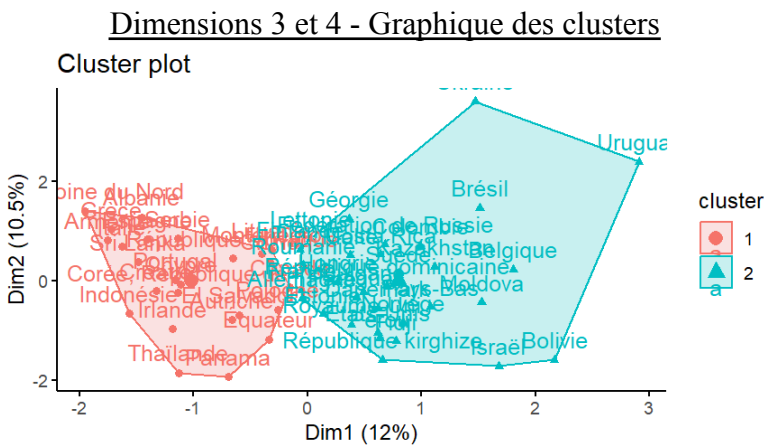


Tableau de résumé de la structure des clusters

Variables	Cluster 1 (valeur moyenne)	Cluster 2 (valeur moyenne)	Test ANOVA (différence significative)

pauvrete	0,32	0,11	Non
pib	9,55	9,66	Non
chomage	1,84	1,63	Non
inflation	1,24	2,81	Oui
indice_gini	34,1	35,05	Non
dep_educ	1,30	1,64	Oui
esp_vie	4,36	4,34	Non
fertilite	0,43	0,60	Oui
pop_rurale	3,55	2,97	Oui
corruption	0,24	0,26	Non

D'après le test de l'ANOVA, nous trouvons qu'il existe des différences significatives entre ces 2 groupes, autrement dit, des p-values inférieures à 0,05 seulement pour 4 variables : l'inflation, les dépenses en éducation, la population rurale et la fertilité.

Puis, les variables impactant le plus la distinction de ces clusters sont, d'après la statistique Eta², ces 4 mêmes variables, avec pour chacune un p-value nettement inférieure à 0,05.

Nous pouvons ainsi séparer nos observations en 2 clusters, en prenant en compte le fait que ces regroupements ont été principalement influencés par les dépenses en éducation ou l'inflation des pays :

- Le premier groupe a donc comme caractéristiques principales des dépenses en éducation plus faible, un taux d'inflation modéré, une population plus rurale et un taux de fertilité faible.
- Le second groupe correspond alors aux pays ayant des dépenses en éducation élevées, ainsi qu'un taux d'inflation et de fertilité élevés, mais avec une population rurale moins importante que le premier groupe.

Nous pouvons alors supposer que le premier cluster représente des pays avec un niveau de développement intermédiaire et des dépenses dans le système éducatif faible alors que le

deuxième cluster représente les pays ayant des dépenses en éducation très importantes, sûrement davantage développées.

Malgré le fait que les groupes formés par les dimensions 3 et 4 de l'ACP sont moins représentatifs de nos données, cela permet quand même de regrouper des pays par rapport à des caractéristiques différentes, ici en se concentrant davantage sur les dépenses en éducation et l'inflation, montrant l'importance de ces variables.

Pour conclure, cette analyse détaillée en cluster donne de nombreuses informations pertinentes. Cela pourrait par exemple permettre, dans le cadre de mise en place de politique publique, d'identifier les points les plus importants sur lesquels il faudrait travailler afin de réduire la pauvreté d'un pays, en spécifiant ces points par rapport au groupe auquel appartient le pays en question.

Après une étude complète de nos données, nous pouvons à présent passer à l'analyse économétrique.

4. Estimation économétrique et interprétation des résultats

A. Régression du modèle de base

1. Qualité du modèle (step)

Afin de choisir le meilleur modèle, en termes de qualité, pour expliquer notre variable dépendante, nous avons utilisé la méthode du step, en déterminant quel modèle possède l'AIC le plus faible.

En premier lieu, nous avons comparé les résultats de différentes méthodes du step : une méthode ascendante (*forward selection*), une méthode descendante (*backward selection*), et une méthode qui combine les deux méthodes précédentes (*stepwise*).

Après avoir réalisé ces trois méthodes, le critère de qualité du modèle le plus faible est le même pour chacune d'entre elles (AIC = 156,27). Ainsi, ce modèle de meilleure qualité est composé de 5 variables explicatives (contre 11 avec le modèle de base) et peut se modéliser comme tel :

$$Pauvrete_i = \beta_0 + \beta_1 \text{espérance de vie}_i + \beta_2 \text{indice de Gini}_i + \beta_3 \text{dépenses en éducation}_i + \beta_4 \text{fertilité}_i + \beta_5 \text{chômage}_i + \varepsilon_i$$

2. Tests statistiques

Les différents tests ont été effectués en émettant l'hypothèse selon laquelle les variables explicatives dans notre modèle sont exogènes. Selon cette hypothèse, chaque variable explicative est indépendante des erreurs du modèle :

$$Cov(X_i, \varepsilon_i) = 0 \quad \forall i$$

En partant de cette supposition, nous pouvons faire plusieurs tests pour évaluer la qualité de notre modèle : le test de normalité des résidus, de la forme fonctionnelle, de la multicolinéarité, de l'homoscédasticité des erreurs, et de la distance de Cooks.

a. Normalité des résidus

Afin de s'assurer de la validité de nos tests et de nos résultats ultérieurs, il est nécessaire, en termes de qualité d'ajustement du modèle, que nos résidus suivent une loi normale pour la suite de notre analyse.

Nous avons alors effectué un test vérifiant la normalité de ces résidus. Étant donné que notre échantillon est de petite taille (57 observations après suppression des valeurs atypiques), nous avons opté pour le test de Shapiro-Wilk qui compare la distribution des données observées avec une distribution normale. Ce test évalue l'hypothèse nulle :

H_0 : Les résidus suivent une loi normale

Suite à ce test, nous avons trouvé une p-value de 0.006078 ($p\text{-value} < 0.05$) ce qui nous permet d'affirmer que l'on ne rejette pas H_0 , c'est à dire que l'on ne rejette pas l'hypothèse selon laquelle les résidus suivent une distribution normale, au seuil de risque de 5%. Par conséquent, la forme fonctionnelle de notre modèle devra donc être modifiée.

b. Forme fonctionnelle

Après cela, nous devons révérifier que la forme fonctionnelle du modèle choisi est appropriée. Nous pouvons vérifier la présence d'erreurs de spécification grâce au test de Ramsey RESET (Regression Equation Specification Error Test), qui nous indique si le modèle sélectionné est linéaire ou non, en partant de l'hypothèse nulle :

H_0 : Le modèle est correctement spécifié

En effectuant ce test, nous trouvons, une p-value de 0.003041 ($p\text{-value} < 0.05$) ; nous rejetons donc l'hypothèse selon laquelle la forme fonctionnelle utilisée est correcte, au seuil de risque de 5%.

Avant de poursuivre notre analyse, nous allons alors devoir modifier notre modèle en ajoutant une forme logarithmique à certaines variables.

Après divers essais, le modèle maximisant le résultat du R^2 est le suivant :

$$Pauvrete_i = \beta_0 + \beta_1 \log(espérance\ de\ vie_i) + \beta_2 \log(indice\ de\ Gini_i) + \beta_3 \log(dépenses\ en\ éducation_i) + \beta_4 \log(fertilité_i) + \beta_5 chômage_i + \varepsilon_i$$

c. Multicolinéarité

Ensuite, il est fondamental que nos variables explicatives ne soient pas fortement corrélées entre elles puisque cela engendrerait un risque de multicolinéarité, rendant difficile le fait d'obtenir une estimation stable.

Nous avons alors utilisé le test de VIF (Variance Inflation Factor) afin de détecter la présence ou non de multicolinéarité. Plus une variable est colinéaire aux autres variables explicatives, plus le coefficient de corrélation multiple au carré est élevé (se rapprochant de 1), plus le VIF augmente à son tour, avec une augmentation du coefficient de cette variable.

Une forte colinéarité – et donc un risque de multicolinéarité, est détectée dès lors que le VIF est supérieur à 5, et devient réellement problématique lorsqu'il dépasse 10.

Dans le cas de notre modèle, chaque variable a un VIF compris entre 1.09 et 1.63 ($VIF < 10$). Par conséquent, nous pouvons conclure qu'il n'y a pas de problème de multicolinéarité entre les variables explicatives de notre modèle, et nous pouvons donc le conserver tel quel.

d. Homoscédasticité des résidus

Afin de s'assurer de la validité des tests statistiques et de l'efficacité de nos estimations, il faut veiller à ce que l'hypothèse d'homoscédasticité des résidus soit vérifiée, c'est-à-dire que la variance des résidus soit constante.

Nous cherchons donc à détecter la présence d'hétéroscédasticité à l'aide du test de Breusch-Pagan. Ce test suit une loi du Chi-deux et a comme hypothèse nulle :

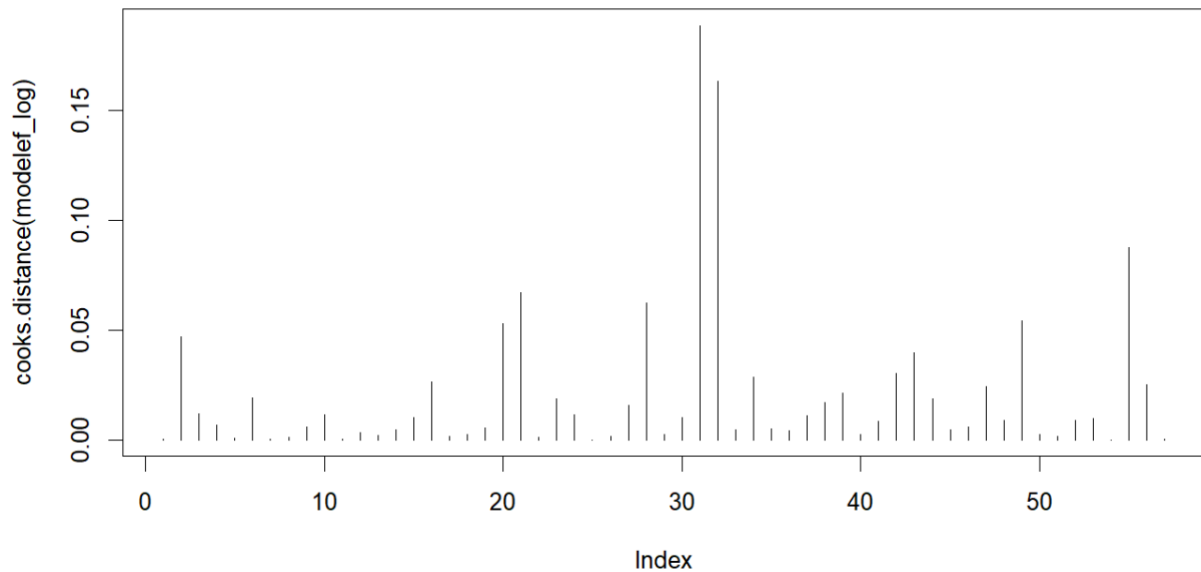
H_0 : Les résidus ont une variance constante (présence d'homoscédasticité)

Avec notre modèle, nous trouvons une p-value de 0.145 ($p\text{-value} > 0.05$), avec laquelle nous ne rejetons pas l'hypothèse d'homoscédasticité des résidus, au seuil de risque de 5%.

e. Distance de Cooks

Enfin, nous cherchons ici à détecter la présence de valeurs extrêmes qui pourraient avoir un impact significatif sur notre régression, et dont la suppression provoquerait une variation drastique de nos résultats. Le graphique de la distance de Cook nous permet de visualiser une potentielle présence de valeurs extrêmes. Appliqué à notre modèle, il se modélise comme tel :

Graphique des distances de Cooks



Ici, nous cherchons à savoir si un ou plusieurs points possèdent une distance de Cook supérieure à 1, avec le nombre d'observations en abscisse et la distance de Cook en ordonnée. Nous pouvons voir qu'aucune des valeurs ne dépasse le seuil de 0.2, et pouvons donc conclure qu'aucun ajustement supplémentaire de notre modèle n'est nécessaire.

Nous nous retrouvons ainsi avec un modèle ajusté correctement spécifié, pour lequel les valeurs influentes sont repérées, les résidus suivent une loi normale et ont une variance constante, et les variables explicatives ne sont pas colinéaires. Nous pouvons alors commencer l'interprétation des résultats de notre régression.

3. Interprétation des résultats

Nous pouvons à présent dresser un tableau récapitulatif des résultats de notre modèle final, réalisé d'après les estimations de la méthode des Moindres Carrés Ordinaires :

Récapitulatif des résultats du modèle final

Résultats de la régression linéaire

Variable	Coefficient	Erreur standard	t-value	p-value	Signif.
(Intercept)	38.2718419	11.4513712	3.342119	0.0015626	**
log(esp_vie)	-11.4779262	2.5577845	-4.487449	0.0000413	***
log(indice_gini)	3.3409179	0.6784874	4.924068	0.0000093	***
log(dep_educ)	-0.9346062	0.3355166	-2.785574	0.0074827	**
log(fertilite)	1.3719797	0.5805309	2.363319	0.0219616	*
chomage	0.1196915	0.0319828	3.742376	0.0004636	***

a. Test de Fisher

Le test de Fisher évalue la significativité globale du modèle estimé. Il compare un modèle avec uniquement une constante à un modèle avec les variables explicatives incluses et mesure l'apport global conjoint de celles-ci. L'hypothèse nulle du test de Fisher est la suivante :

H_0 : Tous les coefficients des variables explicatives sont égaux et nuls

Avec notre modèle final, nous obtenons une p-value de $9.002e-14$ (p-value < 0.05). Nous rejetons donc l'hypothèse H_0 de nullité des coefficients au seuil de risque de 5% et pouvons affirmer qu'il existe au moins une variable dans notre modèle qui explique la variable dépendante.

b. R^2

Le R^2 est un coefficient de détermination qui fait référence à la part de la variation de la variable dépendante expliquée par la somme des variations des variables indépendantes. Il se situe entre 0 et 1 et, plus il est proche de 1, plus les variables explicatives du modèle sont de bonne qualité. Nous pouvons voir que notre R^2 est de 0.739 ce qui signifie qu'environ 74% des variations de la variable dépendante est expliquée par notre modèle. Cela est considéré comme un ratio plutôt élevé qui témoigne d'une bonne qualité de notre modèle.

c. Test de Student

Nous nous intéressons à présent au test de Student qui mesure la significativité des coefficients estimés. Avec ce test, nous cherchons à savoir si une variable explicative a un

impact significatif sur la variable dépendante, pour un seuil de risque donné. L'hypothèse nulle de ce test se modélise comme suit :

H_0 : le coefficient estimé est nul

Pour que cette hypothèse soit rejetée et que donc, le coefficient ait un impact significatif sur la variable à expliquer, il est nécessaire que sa p-value soit inférieure au seuil de risque donné. Dans notre modèle, c'est le cas pour nos 5 variables incluses dans le modèle:

- Pour la variable logarithmique de fertilité, H_0 est rejetée au seuil de risque de 5%
- Pour la variable logarithmique des dépenses en éducation, H_0 est rejetée au seuil de risque de 1%
- Pour les variables logarithmiques de l'espérance de vie et de l'indice de Gini, ainsi que pour la variable du chômage, H_0 est rejetée au seuil de risque de 0,1%

Les variables de notre modèle ont donc un impact significatif sur notre variable dépendante. Nous pouvons à présent en interpréter les coefficients.

d. Interprétation des coefficients

La dernière étape de notre analyse porte sur l'interprétation de nos coefficients significatifs.

Dans un premier temps, une augmentation de 1% de l'espérance de vie réduira d'environ 11,48% la pauvreté.

Ensuite, une augmentation d'1% de l'indice de gini augmentera la pauvreté d'environ 3,34%.

Après cela, nous pouvons voir qu'une augmentation de 1% des dépenses en éducation permettra de diminuer d'environ 0,93% la pauvreté.

En outre, une augmentation de 1% du taux de fertilité sera associée à une augmentation d'environ 1,37% de la pauvreté.

Enfin, une augmentation d'une unité du taux de chômage impactera la pauvreté en l'augmentant d'environ 0,12%.

B. Régression à partir de l'ACP

1. Choix et construction du modèle

Après cela, nous avons également souhaité utiliser les résultats de notre ACP afin d'effectuer une régression linéaire à partir des variables pertinentes ainsi que de variables de contrôle. Cela nous permettra alors de comparer le modèle issu de l'ACP avec le modèle de base.

Dans un premier temps, nous avons effectué une ACP, à partir de nos variables en format logarithmique, en excluant la variable pauvreté (ainsi que les variables qualitatives), afin que cela ait davantage de sens. En effet, puisque la variable pauvreté est la variable à expliquer, l'inclure dans l'ACP risquait de biaiser les résultats de nos coefficients, étant donné que cette variable était très corrélée avec certaines dimensions importantes.

D'après le critère de Kaiser, nous avons alors retenu les 4 premières composantes, en extrayant les coordonnées des individus pour ces dimensions afin de les transformer en variable.

Ensuite, en tant que variables de contrôle, nous avons utilisé les 2 variables qualitatives de notre modèle, "electricite" et "ide", puisque n'étant pas incluses dans l'ACP, cela ne risquait pas d'influencer nos résultats en provoquant une certaine colinéarité. De fait, ces variables détiennent des informations intéressantes ne pouvant pas être perçues par l'ACP. La variable sur l'accès à l'électricité peut notamment montrer des différences structurelles importantes entre les pays observés, tandis que celle des investissements directs étrangers représenterait l'ouverture économique ou la compétitivité d'un pays.

2. Tests statistiques

a. Normalité des résidus

Comme pour le précédent modèle, nous avons commencé par vérifier le test de Shapiro-Wilk. Nous avons alors trouvé une p-value de 0.06378 ($p\text{-value} > 0.05$) ce qui nous permet d'affirmer que l'on accepte H_0 , nous pouvons donc affirmer qu'au seuil de risque de 5%, les résidus suivent une distribution normale.

b. Forme fonctionnelle

Par la suite, nous sommes passés au RESET Test, nous donnant une p-value de 0.004592 (p-value < 0.05) ; nous rejetons ainsi l'hypothèse selon laquelle la forme fonctionnelle utilisée est correcte, au seuil de risque de 5%.

Une modification de notre modèle doit donc s'effectuer. Cependant, il n'est pas possible d'ajouter des formes logarithmiques à nos variables. En effet, les variables quantitatives de ce modèle comportent toutes des valeurs négatives, il ne sera alors pas pertinent de les transformer en logarithme.

Ainsi, après plusieurs essais de différentes formes fonctionnelles, nous avons trouvé une forme étant validée d'après le RESET Test (p-value de 0.05812 > 0.05) :

$$Pauvrete_i = \beta_0 + \beta_1 Dim.1 + \beta_2 Dim.2*electricite + \beta_3 Dim.3*Dim.4 + \beta_4 ide + \varepsilon_i$$

Plusieurs raisons permettent d'expliquer la validité de cette forme.

Pour commencer, la dimension 2 étant fortement corrélée avec l'indice de gini, elle pourrait interagir avec la variable de l'accès à l'électricité, étant donné que cela peut impacter fortement les inégalités de revenus, notamment dans les pays les plus pauvres.

Puis, les dimensions 3 et 4 sont toutes deux grandement corrélées principalement avec les variables du chômage et de l'inflation. Capturer les interactions entre ces composantes pourrait alors être intéressant dans le but d'expliquer la pauvreté d'un pays.

c. Multicolinéarité

Après cela, pour détecter la possible présence multicolinéarité, nous avons utilisé le test de VIF (Variance Inflation Factor).

Pour notre modèle, le VIF le plus élevé est de 2,54. Alors, nous pouvons conclure que nous pouvons garder toutes nos variables explicatives puisqu'il n'y a pas de problème de multicolinéarité entre elles.

d. Homoscédasticité des résidus

Enfin, avec le test de Breusch-Pagan, nous avons cherché à détecter la présence ou non d'hétéroscédasticité.

Nous trouvons une p-value de 0.9544 (p-value > 0.05), montrant que la variance des résidus semble constante à travers les observations, au seuil de risque de 5%.

3. Interprétation des résultats

Désormais, nous pouvons construire un tableau récapitulatif des résultats de notre modèle estimé à partir de l'ACP, réalisé d'après la méthode des MCO :

Tableau récapitulatif des résultats du modèle ACP

Résultats de la régression linéaire

Variable	Coefficient	Erreur standard	t-value	p-value	Signif.
(Intercept)	0.1420831	0.3931964	0.3613539	0.7194207	
Dim.1	-0.6261372	0.0767883	-8.1540736	0.0000000	***
Dim.2	0.3895335	0.1214489	3.2073869	0.0023863	**
electricite100	-0.1726996	0.3225217	-0.5354665	0.5947998	
Dim.3	0.3918441	0.0951162	4.1196358	0.0001490	***
Dim.4	-0.0209742	0.1132941	-0.1851305	0.8539068	
idePositif	0.2743925	0.3445455	0.7963898	0.4297269	
Dim.2:electricite100	0.3103909	0.2113008	1.4689533	0.1483692	
Dim.3:Dim.4	0.2069358	0.0760177	2.7222034	0.0090111	**

a. Test de Fisher

Avec ce modèle, nous obtenons une p-value de 1.853e-15 (p-value < 0.05). Nous rejetons alors l'hypothèse H0 de nullité des coefficients au seuil de risque de 5% ; au moins une variable dans notre modèle explique la variable dépendante.

b. R^2

Notre R^2 est de 0.8214 ce qui signifie qu'environ 82% des variations de la variable dépendante est expliquée par notre modèle, montrant alors une très bonne qualité du modèle.

c. Test de Student

Les variables explicatives de notre modèle ayant un impact significatif sur la variable à expliquer, ayant donc un p-value inférieur à un certain seuil de risque, sont :

- Les variables Dim.1 et Dim.3, sont significatives au seuil de risque de 0.1%
- La variable Dim.2 et l'interaction entre Dim.3 et Dim.4 sont significatives au seuil de risque de 1%

Les autres variables de notre modèle n'ont pas d'impact significatif sur notre variable dépendante.

Nous interpréterons donc uniquement les coefficients des variables ayant un impact significatif dans notre modèle.

d. Interprétation des coefficients

Dans un premier temps, nous pouvons constater que la dimension 1 influencera négativement fortement le niveau de pauvreté d'un pays, puisqu'une augmentation de 1 de la dimension 1 réduira la pauvreté d'environ 63%. Étant très corrélée avec des variables montrant la situation économique du pays, tel que le PIB, nous pouvons en déduire que plus un pays a une économie développée, plus la pauvreté sera faible.

Ensuite, la dimension 2 a un impact positif sur la pauvreté. Cela est sûrement dû au fait que cet axe de l'ACP est corrélé avec la variable de l'indice de gini, représentant les inégalités de revenus. Ainsi, une augmentation d'une unité de la dimension 2 augmentera la pauvreté d'environ 39%.

Après cela, nous constatons qu'une augmentation d'une unité de la dimension 3 augmentera la pauvreté de 39%. Pour une augmentation des interactions entre la dimension 3 et la dimension 4, la pauvreté augmentera alors de 21%. Cela peut s'expliquer par le fait que ces dimensions sont majoritairement corrélées avec les variables représentant le taux de chômage et l'inflation. Logiquement, une augmentation dans ces variables augmenterait le niveau de pauvreté d'un pays.

C. Comparaison des deux modèle

Tableau résumé des résultats des deux modèles

	Modèle classique	Modèle ACP
R ²	0,739	0,8214
AIC	150,0856	134,4646
Test de Fisher	9.002e-14	1.853e-15

Malgré des résultats et des interprétations similaires, le modèle élaboré à partir d'une ACP reste de meilleure qualité, avec un R^2 supérieur et un critère d'AIC plus faible.

Le test de Fisher donnant une p-value inférieure pour le modèle de l'ACP, cela montre également une légère supériorité dans la qualité du modèle.

Néanmoins, ces 2 modèles sont très significatifs et permettent de détecter de façon efficace le niveau de pauvreté d'un pays. En outre, les interprétations pour les estimations faites à partir de ces modèles seront similaires en donnant des informations allant dans le même sens.

5. Modèles Logit / Probit

Dans cette section, nous allons mettre en place un modèle logit et un modèle probit que nous allons comparer afin de trouver la meilleure façon d'expliquer notre phénomène. Pour ce faire, nous utiliserons les mêmes variables que pour l'ACP et la régression linéaire, seulement notre variable à expliquer sera une variable binaire. Nous avons choisi "electricite" comme variable dépendante. Pour rappel, l'évènement qui se produit ici est le fait d'avoir un taux d'accès à l'électricité supérieur à 100%. Nous allons ainsi chercher le meilleur modèle pour prédire la probabilité que cet événement se produise.

A. Modèle Logit

1. Recherche du meilleur modèle

a. Qualité du modèle (step)

En premier lieu, nous avons créé un modèle contenant toutes nos variables explicatives, y compris "pauvrete" qui était autrefois notre variable dépendante, et avons déterminé "electricite" comme variable à expliquer. Ce modèle est de la forme logit, c'est à dire qu'il suit une distribution logistique.

Comme nous l'avons fait précédemment dans le cadre de la régression linéaire, nous avons utilisé la méthode des step afin de choisir le modèle avec la meilleure qualité, c'est-à-dire celui avec l'AIC le plus faible. Après avoir effectué une méthode descendante, une méthode ascendante et une méthode stepwise, l'AIC le plus faible commun aux deux dernières méthodes est de 24.38. Cet AIC correspond au modèle suivant :

$$Electricité_i = \beta_0 + \beta_1 PIB_i + \beta_2 indice\ de\ Gini_i + \beta_3 corruption_i + \beta_4 chômage_i + \varepsilon_i$$

b. Multicolinéarité

Afin de s'assurer qu'il n'y ait pas de risque de multicolinéarité, nous avons effectué un test VIF (Variance Inflation Factor) sur notre modèle logit. Ici, nos variables ont des VIF compris entre 2.07 et 2.58, qui sont donc tous inférieurs à 5, montrant ainsi qu'il n'y a pas de corrélation entre nos différentes variables explicatives.

c. Test de Wald

Nous allons maintenant utiliser le test de Wald qui nous permet de tester la nullité des coefficients, H_0 étant que “tous les coefficients sont nuls”. Nous trouvons une p-value de 3.203376×10^{-24} (largement inférieure à 0.05) ce qui nous permet de rejeter l’hypothèse de nullité de l’ensemble des coefficients. Ainsi, il y a un intérêt à estimer ce modèle au seuil de 5%.

2. Modèle logit final

Après avoir effectué ces tests nous permettant de valider la qualité de notre modèle, nous pouvons en interpréter les résultats.

Tableau récapitulatif des résultats du modèle logit

Résultats de la régression logistique

Variable	Coefficient	Erreur standard	z-value	p-value	Signif.
(Intercept)	13.3193615	6.1086018	2.180427	0.0292258	*
pib	0.0003995	0.0001916	2.085226	0.0370488	*
chomage	0.4958399	0.2753083	1.801035	0.0716973	.
indice_gini	-0.6065121	0.2387517	-2.540346	0.0110743	*
corruption	1.6934729	1.1425197	1.482226	0.1382801	

A noter qu’avec les résultats ci-dessus, nous pouvons seulement interpréter les signes des coefficients obtenus ainsi que leur degré de significativité.

En premier lieu, le PIB et le chômage semblent avoir un impact positif significatif sur la probabilité d’avoir un accès à l’électricité de 100%, aux seuils respectifs de 5% et 10%. Ensuite, ce tableau montre que l’indice de Gini diminue la probabilité d’avoir un accès à l’électricité de 100% au seuil de 5%. Enfin, nous pouvons voir que le niveau de corruption d’un pays ne semble pas avoir d’impact sur le fait d’avoir un accès à l’électricité de 100% (coefficient non significatif).

3. Effets marginaux moyens

C'est avec le calcul des effets marginaux moyens que nous allons pouvoir présenter l'effet moyen de chacune de nos variables explicatives sur notre variable indépendante. Après avoir effectué ces calculs, nous trouvons les résultats suivants :

Tableau des effets marginaux moyens des coefficients du modèle logit

Variable	pib	chomage	indice_gini
Effet marginal moyen	0.00001521827	0.01888717	-0.02310281

- Une augmentation d'un dollar américain du PIB par rapport au PIB moyen entre les pays augmente de 0.000015 la probabilité d'avoir un accès à l'électricité de 100% ;
- Une augmentation d'1% du chômage par rapport au taux de chômage moyen entre les pays augmente de 0.018887 la probabilité d'avoir un accès à l'électricité de 100% ;
- Une diminution d'une unité de l'Indice de Gini par rapport à l'indice moyen entre les pays diminue de 0.023103 la probabilité d'avoir un accès à l'électricité de 100%.

Nous n'avons pas interprété l'effet marginal moyen de la variable "corruption" étant donné que son résultat dans le tableau précédent n'était pas significatif.

B. Modèle Probit

1. Modélisation

Nous pouvons maintenant passer à la modélisation de notre modèle probit qui – à l'inverse du modèle logit, suit une distribution normale. Pour ce faire, nous allons utiliser le même modèle que le précédent, nous évitant ainsi de refaire la méthode de step, le test de multicollinéarité, ainsi que le test de Wald. Pour rappel, le modèle est le suivant :

$$Electricité_i = \beta_0 + \beta_1 PIB_i + \beta_2 indice\ de\ Gini_i + \beta_3 corruption_i + \beta_4 chômage_i + \varepsilon_i$$

Tableau récapitulatif des résultats du modèle probit

Résultats de la régression probit

Variable	Coefficient	Erreur standard	z-value	p-value	Signif.
(Intercept)	7.7144803	3.4309280	2.248511	0.0245436	*
pib	0.0002332	0.0001086	2.148184	0.0316991	*
chomage	0.2841247	0.1511795	1.879386	0.0601918	.
indice_gini	-0.3509146	0.1335132	-2.628313	0.0085809	**
corruption	0.9691826	0.6469936	1.497979	0.1341388	

Comme nous pouvons le voir, les signes des coefficients ainsi que leur seuil de significativité restent identiques à ceux du modèle logit, à part pour l'Indice de Gini, qui est significatif au seuil de 1% (au lieu de 5% avec le modèle logit). Cette légère différence nous sera pertinente lors de la comparaison entre les deux modèles. La variable "corruption" n'est toujours pas significative.

2. Effets marginaux moyens

Comme nous l'avons fait précédemment, nous avons calculé les effets marginaux moyens de notre modèle probit. Nous trouvons :

Tableau des effets marginaux moyens des coefficients du modèle probit

Variable	pib	chomage	indice_gini
Effet marginal moyen	0.0000151396	0.0184450287	-0.0227809446

- Une augmentation d'un dollar américain du PIB par rapport au PIB moyen entre les pays augmente de 0.000015 la probabilité d'avoir un accès à l'électricité de 100% ;
- Une augmentation d'1% du chômage par rapport au taux de chômage moyen entre les pays augmente de 0.018445 la probabilité d'avoir un accès à l'électricité de 100% ;
- Une diminution d'une unité de l'Indice de Gini par rapport à l'indice moyen entre les pays diminue de 0.022781 la probabilité d'avoir un accès à l'électricité de 100%.

C. Comparaison des modèles

Nous allons finir par comparer le modèle logit et le modèle probit afin de déterminer lequel sera le plus adapté pour estimer la probabilité d'avoir un taux d'accès à l'électricité de 100%. Nous allons utiliser plusieurs critères pour comparer ces deux modèles.

En premier lieu, nous utiliserons l'AIC (Akaike Information Criterion) à nouveau ainsi que le BIC (Bayesian Information Criterion), qui est une méthode dérivée de l'AIC qui prend en compte la taille de l'échantillon, et pas seulement le nombre de paramètres. Le meilleur modèle est celui avec l'AIC et le BIC les plus faibles.

Ensuite, nous allons prendre en compte les log-vraisemblances (ou log-likelihood), afin de trouver les valeurs des paramètres qui maximisent la fonction de vraisemblance. Le modèle qui s'ajuste le mieux à nos données est celui avec la log-vraisemblance la plus proche de 0.

Enfin, nous allons utiliser le R^2 de McFadden qui est un pseudo R^2 et qui s'interprète de la même manière qu'un R^2 provenant d'une régression linéaire. Un R^2 de McFadden plus élevé signifie que le modèle a une meilleure qualité d'ajustement.

Tableau récapitulatif des critères de comparaison des deux modèles

	Modèle Logit	Modèle Probit
AIC	24.38422	24.12753
BIC	34.59948	34.34279
Log-Likelihood	-7.192111	-7.063765
Pseudo R^2	0.8017681	0.8053057

Nous pouvons observer que le modèle probit possède un AIC et un BIC légèrement plus faibles, ainsi qu'une log-vraisemblance légèrement plus proche de 0. Le R^2 de McFadden est également plus élevé dans le modèle probit, avec environ 80.53% des variations de la variable dépendante expliqués par ce modèle, indiquant une très bonne qualité d'ajustement. L'ensemble de ces éléments nous permettent de privilégier le modèle probit pour prédire les probabilités du taux d'accès à l'électricité. En outre, les résultats de nos modèles étant assez similaires, un autre élément permettant d'orienter notre choix du meilleur modèle est le seuil

de significativité des coefficients des variables explicatives. En effet, le seuil de 1% de l'indice de Gini au sein du modèle probit montre que celui-ci est plus robuste. Bien que cette information ne constitue pas un critère de comparaison en soit, elle est pertinente car elle permet de renforcer notre décision de choisir le modèle probit, sachant que les critères de comparaison évoqués ci-dessus diffèrent seulement légèrement entre modèles.

Cette dernière étape nous permet de conclure cette section sur les modèles logit/probit ainsi que notre dossier de façon plus générale.

Conclusion

A travers cette étude, en effectuant en amont une analyse de nos données complète, ainsi qu'une analyse de clusters à partir d'une ACP, nous avons alors pu identifier certains facteurs pouvant impacter la pauvreté dans les pays du monde en comparant 2 modèles : l'un classique et l'autre déterminé à partir d'une ACP, avec des régressions effectuées par la méthode des MCO.

Dans un premier temps, l'influence de l'espérance de vie sur la pauvreté montre l'importance de l'investissement dans le domaine de la santé pour un pays, que ce soit en termes d'infrastructure ou de système tels que la sécurité sociale afin de favoriser la santé de la population.

Ensuite, nous pouvons constater qu'un bon système de redistribution des richesses permettrait également de réduire la pauvreté, favorisant notamment l'accès à l'éducation, à la santé ou encore le changement de catégorie socioprofessionnelle, puisque nous pouvons voir que les inégalités économiques représentées par l'indice de Gini impactent positivement le niveau de pauvreté d'un pays.

En outre, l'optimisation du système éducatif d'un pays grâce aux dépenses en éducation, vont aussi permettre de réduire la pauvreté. Cela est notamment dû au fait qu'une population cultivée va pouvoir effectuer un travail nécessitant plus de connaissances, tout en aidant son pays à innover et à être plus concurrentiel par rapport aux autres pays et ainsi, à augmenter leur richesse.

Puis, le nombre de naissances ou le taux de fertilité influencera également la pauvreté d'un pays. De fait, un pays où la population croît rapidement peut rendre compliqué le fait de rendre disponible à chacun un emploi, un logement, ainsi que de la nourriture ou même de l'eau, si les ressources sont limitées, pouvant alors appauvrir la population.

Enfin, l'accès à l'emploi doit être facilité par le gouvernement afin de réduire la pauvreté dans un pays. Cependant, cela peut être compliqué, particulièrement pour les pays pauvres, où moins de d'individus ont accès à l'éducation et où il y aura alors moins de diversité dans le type d'emploi offert et dans les compétences des demandeurs d'emploi.

Une dernière section consacrée à la mise en place de modèles logit et probit pour expliquer la probabilité d'avoir un taux d'accès à l'électricité de 100% nous a permis de mettre en évidence le fait qu'une augmentation du PIB et du taux de chômage, ainsi qu'un indice de Gini plus faible, ont un impact significatif sur la probabilité qu'un pays entier ait accès à l'électricité. Les modèles trouvés parviennent très bien à capturer les variations de notre variable dépendante.

Limites

Malgré de bons résultats global, notre analyse comporte quelques points pouvant être considérés comme des limites.

Pour commencer, l'une des limites principales de notre étude est que les pays faisant partie de l'échantillon ne sont sûrement pas représentatif des pays du monde.

En effet, à cause du manque de données, notre sélection de pays a été limitée à seulement 69 pays. Cependant, nous pouvons supposer que les pays où les données manquent ne sont pas anodins, puisque dans les pays les plus pauvres, avec une situation instable, ou en développement, le recueil de données sera plus compliqué et certainement moins fiable que dans des pays développés.

Après cela, nous avons également supprimé les observations comportant des valeurs atypiques. Le même problème de représentativité se posera donc étant donné que les pays correspondants à ces observations sont certainement des pays avec des cas extrêmes ou des caractéristiques ne correspondant pas à la majorité des pays de notre échantillon. Une autre possibilité est que si l'échantillon d'origine était davantage grand, ces valeurs ne seraient peut-être pas considérées comme atypiques, nous aurions alors pu les garder et ainsi améliorer la qualité et la représentativité de notre modèle.

Les résultats de notre étude ne seront alors pas forcément valides pour tous les pays du globe.

Un autre point important de notre analyse est que nous n'avons pas étudié la question de l'endogénéité de nos variables explicatives. Ceci est dû au fait que l'objectif de cette étude était principalement l'analyse des données, l'élaboration et l'interprétation d'une ACP ainsi que d'un modèle logit/probit. De plus, les variables que nous avons sélectionnées étant des variables pouvant potentiellement être influencées par divers autres facteurs, il y avait alors beaucoup de possibilités d'endogénéité, rendant les tests à effectuer plus compliqués. Nous avons décidé de ne pas le prendre en compte et de partir du principe que nos variables sont exogènes, c'est-à-dire indépendantes du mouvement interne du modèle, tout en expliquant et impactant celui-ci.

Enfin, concernant les modèles logit et probit, nous avons uniquement calculé les effets marginaux moyens et non les effets marginaux individuels, afin d'avoir une vue d'ensemble

de l'effet des variables, ainsi que pour mieux comparer nos modèles. Par conséquent, les variations d'effets entre les observations n'ont pas pu être prises en compte. Il est également possible que les effets marginaux moyens ne soient pas représentatifs si les effets marginaux varient beaucoup dans l'échantillon.

Bibliographie

Akbar, M., Khan, M., Farooqe, H., & Kaleemullah. (2019). *Public spending, education and poverty: A cross-country analysis*. Journal of Multidisciplinary Approaches in Science, 4(1), 12-20. Journal of Multidisciplinary Approaches in Science (JMAS).

Eshun, J., & Baah, J. A. (2019). *What does recent survey data say about the effect of corruption on poverty in Africa?* Journal of Economics, Management and Trade, 25(6), 1-13.

Harvey, P.-A., & Nguyen, M. (2020). *L'inégalité face à l'inflation: Une croissance du coût de la vie selon les revenus des familles*. IRIS (Institut de recherche et d'informations socioéconomiques)

Henneh, O. F. (2023). *Impact of economic growth on poverty reduction: Evidence from Ghana*. Texila International Journal of Management, 9(2). <https://doi.org/10.21522/TIJMG.2015.09.02.Art007>

Johan, M., Febrianti, L. A., Wahyuningtyas, I., Salim, F., & Prawira, F. D. (2024). *The influence of unemployment on poverty levels in Indonesia*. ORGANIZE: Journal of Economics, Management and Finance, 3(2), 91–100. <https://doi.org/10.58355/organize.v3i2.65>

Kamdem Kamdem, M. (2012). *La contribution de l'énergie à la réduction de la pauvreté en milieu rural au Cameroun*. ICBE-RF Research Report No. 08/12. Investment Climate and Business Environment Research Fund (ICBE-RF), Université de Yaoundé II, Cameroun.

Khan, M. H. (2002). *La pauvreté rurale dans les pays en développement: Orientation pour l'action publique*. Economic Issues Series, No. 26. International Monetary Fund.

Lakner, C., Mahler, D. G., Negre, M., & Prydz, E. B. (2020). *How much does reducing inequality matter for global poverty?* World Bank Group.

Moujahid, M., & Khariss, M. (2021). *Investissement direct étranger et réduction de la pauvreté : Essai d'élaboration d'une revue de la littérature*. International Journal of Accounting, Finance, Auditing, Management and Economics, 2(4), 356-373. <https://doi.org/10.5281/zenodo.5138765>

Sudaryati, S., Ahmad, A. A., & Suprpto. (2021). *The effect of average length of schooling, life expectancy, and economic growth on poverty in Banjarnegara Regency 2005-2019*.

Eko-Regional: Jurnal Pembangunan Ekonomi Wilayah, 16(1), 29-35. Faculty of Economics and Business, Universitas Jenderal Soedirman.

Talukdar, S. R. (2012). *The effect of inflation on poverty in developing countries: A panel data analysis* (Master of Arts thesis, Texas Tech University). 1-56.

Wietzke, F.-B. (2020). *Poverty, inequality, and fertility: The contribution of demographic change to global poverty reduction*. Population and Development Review, 46(1), 65-99. <https://doi.org/10.1111/padr.12317>

World Bank. (2024). *Poverty overview*.

Annexe

Annexe 1 : Tableau analyse descriptive - variables quantitatives

	variable <chr>	moyenne <dbl>	médiane <dbl>	minimum <dbl>	maximum <dbl>	`écart-type` <dbl>
1	pauvrete	9.77	1.3	0.1	89.1	19.4
2	pib	20898.	10144.	512.	112726.	22856.
3	chomage	6.60	5.55	0.716	17.4	3.63
4	inflation	6.82	2.10	-0.355	255.	30.8
5	indice_gini	35.5	34.2	23.2	53.5	6.94
6	dep_educ	4.52	4.25	0.964	8.09	1.45
7	esp_vie	76.4	76.8	61.2	83.8	5.62
8	fertilite	2.00	1.7	0.918	4.81	0.859
9	pop_rurale	33.8	31.9	1.96	82.8	18.9
10	corruption	0.169	0.00690	-1.56	2.12	1.06

Annexe 2 : Valeurs atypiques - pop_rurale

Number of Outliers Detected: 0

	i	Mean.i	SD.i	Value	Obs.Num	R.i+1	lambda.i+1	Outlier
1	0	33.82828	18.90050	82.826	48	2.592404	3.252277	FALSE
2	1	33.10772	18.06104	81.415	38	2.674668	3.246863	FALSE

Annexe 3 : Valeurs atypiques - fertilite

Number of Outliers Detected: 6

	i	Mean.i	SD.i	Value	Obs.Num	R.i+1	lambda.i+1	Outlier
1	0	1.995159	0.8591069	4.809	64	3.275309	3.252277	TRUE
2	1	1.953779	0.7932012	4.783	46	3.566839	3.246863	TRUE
3	2	1.911552	0.7180603	4.521	47	3.634023	3.241349	TRUE
4	3	1.872015	0.6459092	4.072	48	3.406028	3.235733	TRUE
5	4	1.838169	0.5890077	3.813	68	3.352810	3.230010	TRUE
6	5	1.807313	0.5381129	3.599	69	3.329575	3.224177	TRUE
7	6	1.778873	0.4915664	3.300	36	3.094448	3.218230	FALSE

Annexe 4 : Valeurs atypiques - dep_educ

Number of Outliers Detected: 0

	i	Mean.i	SD.i	Value	Obs.Num	R.i+1	lambda.i+1	Outlier
1	0	4.517434	1.452982	8.0910397	6	2.459498	3.252277	FALSE
2	1	4.464881	1.396157	0.9643991	29	2.507226	3.246863	FALSE
3	2	4.517127	1.338044	7.8645601	50	2.501736	3.241349	FALSE

Annexe 5 : Valeurs atypiques - indice_gini

Number of Outliers Detected: 0

	i	Mean.i	SD.i	value	Obs.Num	R.i+1	lambda.i+1	outlier
1	0	35.48841	6.938078	53.5	7	2.596050	3.252277	FALSE
2	1	35.22353	6.628886	51.3	10	2.425215	3.246863	FALSE

Annexe 6 : Valeurs atypiques - inflation

Number of Outliers Detected: 5

	i	Mean.i	SD.i	value	Obs.Num	R.i+1	lambda.i+1	outlier
1	0	6.816812	30.787085	255.304991	69	8.071182	3.252277	TRUE
2	1	3.162574	5.181802	39.907346	32	7.091118	3.246863	TRUE
3	2	2.614145	2.548594	15.176822	63	4.929258	3.241349	TRUE
4	3	2.423801	2.032353	9.370836	48	3.418223	3.235733	TRUE
5	4	2.316924	1.851840	9.152800	17	3.691396	3.230010	TRUE

Annexe 7 : Valeurs atypiques - pib

Number of Outliers Detected: 1

	i	Mean.i	SD.i	value	Obs.Num	R.i+1	lambda.i+1	outlier
1	0	20897.73	22856.10	112726.44	40	4.017690	3.252277	TRUE
2	1	19547.31	20062.13	80848.30	31	3.055558	3.246863	FALSE
3	2	18632.37	18729.68	76430.59	50	3.085916	3.241349	FALSE

Annexe 8 : Valeurs atypiques - pauvreté

Number of Outliers Detected: 7

	i	Mean.i	SD.i	value	Obs.Num	R.i+1	lambda.i+1	outlier
1	0	9.773913	19.385733	89.1	48	4.091983	3.252277	TRUE
2	1	8.607353	16.914926	88.6	46	4.729116	3.246863	TRUE
3	2	7.413433	13.858242	71.8	64	4.646085	3.241349	TRUE
4	3	6.437879	11.413161	64.5	69	5.087296	3.235733	TRUE
5	4	5.544615	8.877500	44.0	30	4.331781	3.230010	TRUE
6	5	4.943750	7.498061	34.9	68	3.995200	3.224177	TRUE
7	6	4.468254	6.513381	26.4	26	3.367183	3.218230	TRUE

Annexe 9 : Valeurs atypiques - chômage

Number of Outliers Detected: 0

	i	Mean.i	SD.i	value	Obs.Num	R.i+1	lambda.i+1	outlier
1	0	6.603464	3.626863	17.414	43	2.980685	3.252277	FALSE
2	1	6.444485	3.403017	17.045	25	3.115035	3.246863	FALSE
3	2	6.286269	3.166699	15.128	45	2.792097	3.241349	FALSE

Annexe 10 : Coefficients de symétrie et d'aplatissement des variables quantitatives

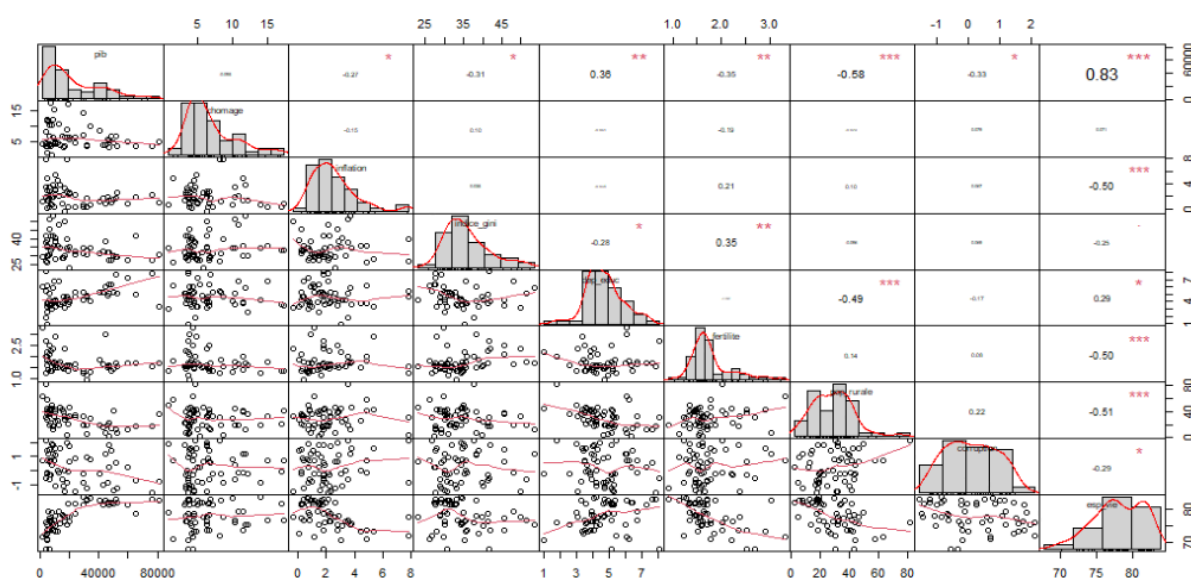
	Variable	Skewness	Kurtosis
1	pauvrete	1.9654313	3.9175448
2	pib	1.1278192	0.4190371
3	chomage	1.2794762	1.1695807
4	inflation	1.5021374	2.9923405
5	indice_gini	1.0155933	0.7326944
6	dep_educ	0.2608465	0.4966326
7	esp_vie	-0.4418703	-0.5472996
8	fertilite	1.3972933	1.9348153
9	pop_rurale	0.6390786	1.2008704
10	corruption	0.0620253	-1.1320146

Annexe 11 : Coefficients de corrélation de Spearman

	pib	chomage	inflation	indice_gini	dep_educ
pib	1.00000000	-0.09460226	-0.27255639	-0.31390600	0.360109547
chomage	-0.09460226	1.00000000	-0.15096174	0.10100650	-0.061353471
inflation	-0.27255639	-0.15096174	1.00000000	0.03468396	-0.034776127
indice_gini	-0.31390600	0.10100650	0.03468396	1.00000000	-0.278976290
dep_educ	0.36010955	-0.06135347	-0.03477613	-0.27897629	1.000000000
fertilite	-0.34878294	-0.19453853	0.20805108	0.34637080	-0.008524707
pop_rurale	-0.57823438	-0.02933027	0.09975369	-0.09646678	-0.492699607
corruption	-0.33415999	0.07939184	0.06703947	0.06933839	-0.167169167
esp_vie	0.82901219	0.07113805	-0.49714804	-0.25316046	0.286019868

	fertilite	pop_rurale	corruption	esp_vie
pib	-0.348782940	-0.57823438	-0.33415999	0.82901219
chomage	-0.194538531	-0.02933027	0.07939184	0.07113805
inflation	0.208051082	0.09975369	0.06703947	-0.49714804
indice_gini	0.346370799	-0.09646678	0.06933839	-0.25316046
dep_educ	-0.008524707	-0.49269961	-0.16716917	0.28601987
fertilite	1.000000000	0.14011603	0.07957659	-0.50478074
pop_rurale	0.140116035	1.00000000	0.21943432	-0.50557428
corruption	0.079576587	0.21943432	1.00000000	-0.28705731
esp_vie	-0.504780737	-0.50557428	-0.28705731	1.00000000

Annexe 12 : Graphique des coefficients de corrélation de Spearman



Annexe 13 : Test de corrélation de Pearson entre esp_vie et pib

Pearson's product-moment correlation

```
data: pauvrete2$esp_vie and pauvrete2$pib
t = 8.2695, df = 55, p-value = 3.179e-11
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.6003673 0.8417524
sample estimates:
      cor
0.7444724
```

Annexe 14 : Tableau de contingence des variables qualitatives

	Négatif	Positif
<100	1	18
100	5	33

Annexe 15 : Test de Khi-deux

Pearson's Chi-squared test with Yates' continuity correction

```
data: mytable
X-squared = 0.20956, df = 1, p-value = 0.6471
```

Annexe 16 : Test de comparaison des moyennes - electricite et pauvrete

welch Two Sample t-test

```
data: pauvrete by electricite
t = 3.9005, df = 24.642, p-value = 0.0006525
alternative hypothesis: true difference in means between group <100 and group 100 is not equal to 0
95 percent confidence interval:
 2.654543 8.603352
sample estimates:
mean in group <100 mean in group 100
 7.331579          1.702632
```

Annexe 17 : Test de comparaison des moyennes - ide et pauvrete

welch Two Sample t-test

```
data: pauvrete by ide
t = -2.2994, df = 13.119, p-value = 0.03853
alternative hypothesis: true difference in means between group Négatif and group Positif is not equal to 0
95 percent confidence interval:
 -5.2267786 -0.1653783
sample estimates:
mean in group Négatif mean in group Positif
 1.166667             3.862745
```

Annexe 18 : Statistique éta carré - electricite


```

> etaSquared(aov(pauvrete~electricite, pauvrete2))[1]
[1] 0.2767365
> etaSquared(aov(pib~electricite, pauvrete2))[1]
[1] 0.2757229
> etaSquared(aov(chomage~electricite, pauvrete2))[1]
[1] 0.0116405
> etaSquared(aov(inflation~electricite, pauvrete2))[1]
[1] 0.0129906
> etaSquared(aov(indice_gini~electricite, pauvrete2))[1]
[1] 0.4508027
> etaSquared(aov(dep_educ~electricite, pauvrete2))[1]
[1] 0.04368557
> etaSquared(aov(esp_vie~electricite, pauvrete2))[1]
[1] 0.3016324
> etaSquared(aov(fertilite~electricite, pauvrete2))[1]
[1] 0.1957199
> etaSquared(aov(pop_rurale~electricite, pauvrete2))[1]
[1] 0.02831495
> etaSquared(aov(corruption~electricite, pauvrete2))[1]

```

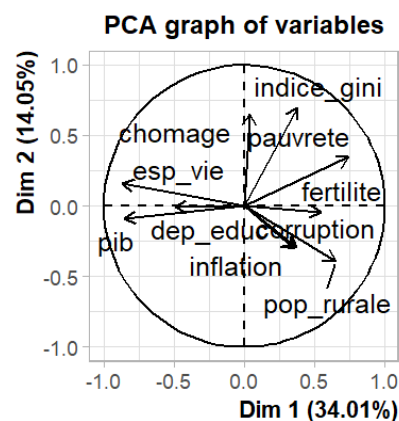
Annexe 19 : Statistique éta carré - ide

```

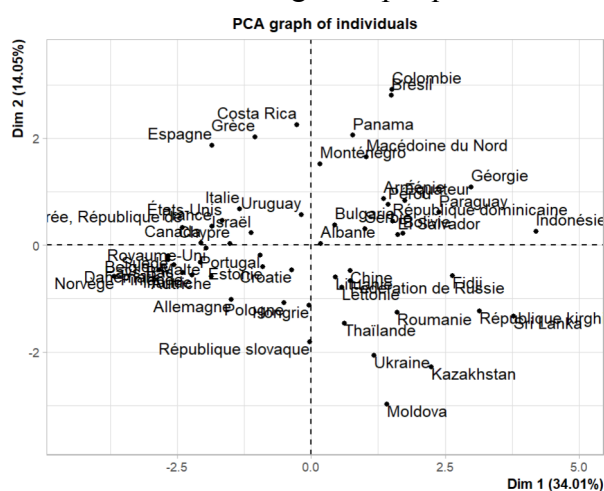
> etaSquared(aov(pauvrete~ide, pauvrete2))[1]
[1] 0.02690676
> etaSquared(aov(pib~ide, pauvrete2))[1]
[1] 0.196417
> etaSquared(aov(chomage~ide, pauvrete2))[1]
[1] 0.03465928
> etaSquared(aov(inflation~ide, pauvrete2))[1]
[1] 0.01467619
> etaSquared(aov(indice_gini~ide, pauvrete2))[1]
[1] 0.033738
> etaSquared(aov(dep_educ~ide, pauvrete2))[1]
[1] 0.08668225
> etaSquared(aov(esp_vie~ide, pauvrete2))[1]
[1] 0.02403709
> etaSquared(aov(fertilite~ide, pauvrete2))[1]
[1] 0.0009870378
> etaSquared(aov(pop_rurale~ide, pauvrete2))[1]
[1] 0.03404313
> etaSquared(aov(corruption~ide, pauvrete2))[1]
[1] 0.07340518

```

Annexe 20 : ACP sans log - Graphique des variables



Annexe 21 : ACP sans log - Graphique des individus



Annexe 22 : ACP sans log - Valeurs propres

	eigenvalue	percentage of variance	cumulative percentage of variance
comp 1	3.4010543	34.010543	34.01054
comp 2	1.4045324	14.045324	48.05587
comp 3	1.3376828	13.376828	61.43270
comp 4	1.0666623	10.666623	72.09932
comp 5	0.8933509	8.933509	81.03283
comp 6	0.7236381	7.236381	88.26921
comp 7	0.4391151	4.391151	92.66036
comp 8	0.3321031	3.321031	95.98139
comp 9	0.2762100	2.762100	98.74349
comp 10	0.1256509	1.256509	100.00000

Annexe 23 : ACP sans log - Contribution des variables

	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
pauvrete	16.21764373	8.518564557	1.1146624	2.9151833	3.1761411
pib	21.52664851	0.609586840	3.0380450	2.6975409	0.3107951
chomage	0.04484391	30.357127311	18.6749972	6.6675651	1.1089278
inflation	3.66254833	6.569846623	0.9839694	54.3221773	12.0899352
indice_gini	4.19559814	34.843366579	5.8044886	0.1980621	1.6463639
dep_educ	7.13397698	0.005715944	22.6181930	8.2765943	9.2274978
esp_vie	22.21188804	1.752804568	3.5668431	2.8227512	0.5698290
fertilite	8.67112392	0.125843556	30.6247553	7.1029041	5.0083629
pop_rurale	12.32314970	11.069600337	7.6144084	12.0016032	0.6477550
corruption	4.01257876	6.147543684	5.9596377	2.9956186	66.2143922

Annexe 24 : ACP sans log - Corrélation des variables

	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
pauvrete	0.74267818	0.345898830	0.1221092	-0.17633820	-0.16844609
pib	-0.85564772	-0.092530237	0.2015922	-0.16962798	-0.05269242
chomage	0.03905337	0.652974497	-0.4998122	0.26668409	0.09953199
inflation	0.35293804	-0.303769033	0.1147275	0.76120574	-0.32864198
indice_gini	0.37774935	0.699561561	0.2786497	0.04596362	0.12127575
dep_educ	-0.49257531	-0.008960038	0.5500543	0.29712508	0.28711311
esp_vie	-0.86915958	0.156903500	-0.2184332	-0.17352009	0.07134825
fertilite	0.54305583	-0.042041807	0.6400485	-0.27525261	0.21152366
pop_rurale	0.64739248	-0.394304609	-0.3191499	-0.35779404	-0.07607053
corruption	0.36941844	-0.293843911	-0.2823492	0.17875439	0.76910785

Annexe 25 : ACP log - Valeurs propres

	eigenvalue	percentage of variance	cumulative percentage of variance
comp 1	3.6575705	36.575705	36.57570
comp 2	1.4357430	14.357430	50.93313
comp 3	1.1981307	11.981307	62.91444
comp 4	1.0537207	10.537207	73.45165
comp 5	0.8757507	8.757507	82.20916
comp 6	0.7336615	7.336615	89.54577
comp 7	0.4824976	4.824976	94.37075
comp 8	0.3155959	3.155959	97.52671
comp 9	0.1465399	1.465399	98.99210
comp 10	0.1007896	1.007896	100.00000

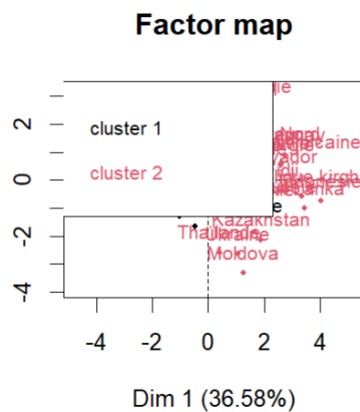
Annexe 26 : ACP log - Contribution des variables

	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
pauvrete	19.659924320	11.5955813	0.67364199	0.14938376	0.06853908
pib	22.676616847	0.4679599	0.00768558	1.68460717	0.01001047
chomage	0.009700721	24.6617242	4.27759108	39.65886576	0.01149090
inflation	2.414853838	8.1803953	25.79760922	27.57153977	12.30795289
indice_gini	5.292592961	34.5463306	2.05001682	0.59616644	0.78452563
dep_educ	6.525599560	1.3138604	26.92682461	0.10003338	5.41205199
esp_vie	20.709998519	2.9663291	5.58536164	0.03562481	1.74427690
fertilite	9.837032270	0.8572693	12.20280034	21.14494654	1.97968973
pop_rurale	9.796960165	7.6570401	22.36982179	1.31922232	1.39178258
corruption	3.076720798	7.7535097	0.10864692	7.73961005	76.28967984

Annexe 27 : ACP log - Corrélation des variables

	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
pauvrete	0.84798325	0.40802297	-0.089839366	-0.03967477	-0.024499622
pib	-0.91072128	0.08196768	0.009596004	-0.13323308	-0.009363051
chomage	0.01883642	0.59504535	-0.226387132	0.64644696	0.010031532
inflation	0.29719519	-0.34270899	0.555957809	0.53900559	-0.328309277
indice_gini	0.43997763	0.70427020	0.156722307	-0.07925862	0.082888410
dep_educ	-0.48854724	0.13734503	0.567995213	-0.03246648	0.217706413
esp_vie	-0.87033487	0.20637069	-0.258688876	-0.01937488	0.123594162
fertilite	0.59983030	0.11094226	0.382368280	-0.47202614	0.131670598
pop_rurale	0.59860732	-0.33156510	-0.517706201	-0.11790216	-0.110401746
corruption	0.33545973	-0.33364723	-0.036079525	0.28557639	0.817378364

Annexe 28 : Dimensions 1 et 2 - Factor Map



Annexe 29 : Dimensions 1 et 2 - Résumé des clusters

	pauvrete	pib	chomage	inflation	indice_gini	dep_educ	esp_vie	fertilite
1	-0.9876582	10.467351	1.690954	1.648730	31.43333	1.616372	4.39646	0.4148536
2	1.2763909	8.846559	1.745114	2.545331	37.51667	1.380108	4.31319	0.6237299
	pop_rurale	corruption	clust					
1	2.883891	-0.2033832	1					
2	3.532244	0.6632859	2					

Annexe 30 : Dimensions 1 et 2 - Résultats Eta²

	Eta2	P-value
pib	0.70875932	2.335543e-16
esp_vie	0.59337700	2.459345e-12
pauvrete	0.52336259	2.059995e-10
pop_rurale	0.23405545	1.375248e-04
indice_gini	0.21372790	2.940983e-04
fertilite	0.18006061	1.003173e-03
corruption	0.16693458	1.603385e-03
dep_educ	0.11439461	1.007402e-02
inflation	0.06995988	4.678407e-02

Annexe 31 : Dimensions 1 et 2 - Résultats ANOVA

\$pauvrete

One-way analysis of means

data: pauvrete and clust

F = 60.392, num df = 1, denom df = 55, p-value = 2.06e-10

\$pib

One-way analysis of means

data: pib and clust

F = 133.85, num df = 1, denom df = 55, p-value = 2.336e-16

\$chomage

One-way analysis of means

data: chomage and clust

F = 0.12447, num df = 1, denom df = 55, p-value = 0.7256

\$inflation

One-way analysis of means

data: inflation and clust

F = 4.1372, num df = 1, denom df = 55, p-value = 0.04678

\$indice_gini

One-way analysis of means

data: indice_gini and clust

F = 14.95, num df = 1, denom df = 55, p-value = 0.0002941

\$dep_educ

One-way analysis of means

data: dep_educ and clust

F = 7.1044, num df = 1, denom df = 55, p-value = 0.01007

\$esp_vie

One-way analysis of means

data: esp_vie and clust

F = 80.26, num df = 1, denom df = 55, p-value = 2.459e-12

\$fertilite

One-way analysis of means

data: fertilite and clust

F = 12.078, num df = 1, denom df = 55, p-value = 0.001003

\$pop_rurale

One-way analysis of means

data: pop_rurale and clust

F = 16.807, num df = 1, denom df = 55, p-value = 0.0001375

\$corruption

One-way analysis of means

data: corruption and clust

F = 11.021, num df = 1, denom df = 55, p-value = 0.001603

Annexe 32 : Dimensions 3 et 4 - Résumé des clusters

	pauvrete	pib	chomage	inflation	indice_gini	dep_educ	esp_vie	fertilite
1	0.3216464	9.548754	1.836861	1.237387	34.09600	1.303831	4.364689	0.4308024
2	0.1119936	9.665513	1.627739	2.810656	35.05625	1.639047	4.343216	0.5982151

	pop_rurale	corruption	clust
1	3.548028	0.2400674	1
2	2.972865	0.2626733	2

Annexe 33 : Dimensions 3 et 4 - Résultats Eta²

	Eta2	P-value
dep_educ	0.2274402	0.0001764247
inflation	0.2127464	0.0003049746
pop_rurale	0.1819204	0.0009383129
fertilite	0.1142411	0.0101276144

Annexe 34 : Dimensions 3 et 4 - Résultats ANOVA

\$pauvrete

One-way analysis of means

data: pauvrete and clust

F = 0.24487, num df = 1, denom df = 55, p-value = 0.6227

\$pib

One-way analysis of means

data: pib and clust

F = 0.20053, num df = 1, denom df = 55, p-value = 0.6561

\$chomage

One-way analysis of means

data: chomage and clust

F = 1.8914, num df = 1, denom df = 55, p-value = 0.1746

\$inflation

One-way analysis of means

data: inflation and clust

F = 14.863, num df = 1, denom df = 55, p-value = 0.000305

\$indice_gini

One-way analysis of means

data: indice_gini and clust

F = 0.29081, num df = 1, denom df = 55, p-value = 0.5919

\$dep_educ

One-way analysis of means

data: dep_educ and clust

F = 16.192, num df = 1, denom df = 55, p-value = 0.0001764

\$esp_vie

One-way analysis of means

data: esp_vie and clust

F = 2.2303, num df = 1, denom df = 55, p-value = 0.141

\$fertilite

One-way analysis of means

data: fertilite and clust

F = 7.0936, num df = 1, denom df = 55, p-value = 0.01013

\$pop_rurale

One-way analysis of means

data: pop_rurale and clust

F = 12.231, num df = 1, denom df = 55, p-value = 0.0009383

\$corruption

One-way analysis of means

data: corruption and clust

F = 0.0061702, num df = 1, denom df = 55, p-value = 0.9377

Annexe 35 : Test de Shapiro-Wilk - modelef

shapiro-wilk normality test

data: residus

W = 0.93844, p-value = 0.006078

Annexe 36 : RESET Test - modelef

RESET test

data: modelef

RESET = 6.5385, df1 = 2, df2 = 49, p-value = 0.003041

Annexe 37 : RESET Test après log - modelef_log

RESET test

data: modelef_log

RESET = 2.37, df1 = 2, df2 = 49, p-value = 0.1041

Annexe 38 : VIF Test - modelef_log

log(esp_vie)	log(indice_gini)	log(dep_educ)	log(fertilite)
1.525072	1.188124	1.095804	1.629020
chomage			
1.123762			

Annexe 39 : Test de Breusch-Pagan - modelef_log

studentized Breusch-Pagan test

data: modelef_log
BP = 8.2114, df = 5, p-value = 0.145

Annexe 40 : Résultats de la régression -modelef_log

Call:
lm(formula = log(pauvrete) ~ log(esp_vie) + log(indice_gini) +
log(dep_educ) + log(fertilite) + chomage, data = pauvrete2)

Residuals:

Min	1Q	Median	3Q	Max
-1.7281	-0.5868	0.1887	0.5885	1.4205

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	38.27184	11.45137	3.342	0.001563	**
log(esp_vie)	-11.47793	2.55778	-4.487	4.13e-05	***
log(indice_gini)	3.34092	0.67849	4.924	9.31e-06	***
log(dep_educ)	-0.93461	0.33552	-2.786	0.007483	**
log(fertilite)	1.37198	0.58053	2.363	0.021962	*
chomage	0.11969	0.03198	3.742	0.000464	***

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

Residual standard error: 0.844 on 51 degrees of freedom

Multiple R-squared: 0.739, Adjusted R-squared: 0.7134

F-statistic: 28.88 on 5 and 51 DF, p-value: 9.002e-14

Annexe 41 : Valeurs propres - ACP sans pauvrete

	eigenvalue	percentage of variance	cumulative percentage of variance
comp 1	3.0302961	33.669957	33.66996
comp 2	1.2585287	13.983652	47.65361
comp 3	1.1623099	12.914555	60.56816
comp 4	1.0493751	11.659723	72.22789
comp 5	0.8746815	9.718684	81.94657
comp 6	0.7335648	8.150720	90.09729
comp 7	0.4811384	5.345982	95.44327
comp 8	0.3032074	3.368972	98.81224
comp 9	0.1068980	1.187756	100.00000

Annexe 42 : Contribution des variables - ACP sans pauvrete

	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
pib	27.1324311	0.1868907	0.5977623	1.155607958	5.876454e-04
chomage	0.2217916	14.6476147	29.0953998	29.028803275	6.313236e-03
inflation	4.7651294	0.5568799	25.6523293	34.579885036	1.099450e+01
indice_gini	3.3369635	42.9018023	6.8156547	2.826696547	3.101399e-01
dep_educ	7.8852247	9.5453387	19.3464403	0.005735262	6.005609e+00
esp_vie	26.8360726	0.5075050	4.1507451	0.085162138	1.799028e+00
fertilite	11.2883556	10.1569991	4.7777411	22.330964438	1.563087e+00
pop_rurale	13.2673322	17.8831694	9.4304082	1.759901163	1.560861e+00
corruption	5.2666993	3.6138001	0.1335194	8.227244182	7.775987e+01

Annexe 43 : Corrélation des variables - ACP sans pauvrete

	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
pib	0.90674859	-0.04849818	-0.08335376	-0.110121124	-0.002267163
chomage	0.08198135	0.42935351	0.58153135	0.551924838	-0.007431064
inflation	-0.37999675	0.08371675	-0.54603989	0.602389156	-0.310107857
indice_gini	-0.31799352	0.73480032	0.28145875	-0.172228479	0.052083934
dep_educ	0.48882068	0.34659894	-0.47419995	0.007757861	0.229194148
esp_vie	0.90178294	-0.07991931	0.21964636	-0.029894318	0.125442284
fertilite	-0.58486802	0.35753147	-0.23565262	-0.484082199	0.116927481
pop_rurale	-0.63406581	-0.47440997	0.33107487	-0.135896888	-0.116844201
corruption	-0.39949541	-0.21326207	-0.03939427	0.293827586	0.824712822

Annexe 44 : Test de Shapiro - Wilk - modele_acp

shapiro-wilk normality test

data: residus

W = 0.96105, p-value = 0.06378

Annexe 45 : RESET Test - modele_acp

RESET test

data: modele_acp

RESET = 6.0351, df1 = 2, df2 = 48, p-value = 0.004592

Annexe 46 : RESET Test après changement de la forme fonctionnelle - modele_acp_2

RESET test

data: modele_acp_2

RESET = 3.0287, df1 = 2, df2 = 46, p-value = 0.05812

Annexe 47 : VIF Test - modele_acp_2

Dim.1	Dim.2	electricite	Dim.3
1.966354	2.042853	2.543856	1.157225
Dim.4	ide	Dim.2:electricite	Dim.3:Dim.4
1.482286	1.230416	1.725754	1.359471

Annexe 48 : Test de Breusch-Pagan - modele_acp_2

studentized Breusch-Pagan test

data: modele_acp_2

BP = 2.65, df = 8, p-value = 0.9544

Annexe 49 : Résultats de la régression - modèle_acp_2

```
Call:
lm(formula = pauvrete ~ Dim.1 + Dim.2 * electricite + Dim.3 *
    Dim.4 + ide, data = dimensions)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-1.80886	-0.44436	0.08189	0.43359	1.28594

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	0.14208	0.39320	0.361	0.719421	
Dim.1	-0.62614	0.07679	-8.154	1.29e-10	***
Dim.2	0.38953	0.12145	3.207	0.002386	**
electricite100	-0.17270	0.32252	-0.535	0.594800	
Dim.3	0.39184	0.09512	4.120	0.000149	***
Dim.4	-0.02097	0.11329	-0.185	0.853907	
idePositif	0.27439	0.34455	0.796	0.429727	
Dim.2:electricite100	0.31039	0.21130	1.469	0.148369	
Dim.3:Dim.4	0.20694	0.07602	2.722	0.009011	**

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

Residual standard error: 0.7197 on 48 degrees of freedom
 Multiple R-squared: 0.8214, Adjusted R-squared: 0.7916
 F-statistic: 27.59 on 8 and 48 DF, p-value: 1.853e-15

Annexe 50 : VIF Test - modelef_logit

pib	chomage	indice_gini	corruption
2.271633	2.065679	2.294541	2.575445

Annexe 51 : Wald Test - modelef_logit

```
> chi2<-2*(modelef_logit$null.deviance-modelef_logit$deviance)
> print(chi2)
[1] 116.3568
> ddl<-modelef_logit$df.null-modelef_logit$df.residual
> print(ddl)
[1] 4
> pvalue<-pchisq(chi2,ddl,lower.tail=F)
> print(pvalue)
[1] 3.203376e-24
```

Annexe 52 : Résultats de la régression - modelef_logit

```
call:
glm(formula = electricite ~ pib + chomage + indice_gini + corruption,
     family = binomial(link = "logit"), data = pauvrete2, x = TRUE)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	13.3193615	6.1086018	2.180	0.0292	*
pib	0.0003995	0.0001916	2.085	0.0370	*
chomage	0.4958399	0.2753083	1.801	0.0717	.
indice_gini	-0.6065121	0.2387517	-2.540	0.0111	*
corruption	1.6934729	1.1425197	1.482	0.1383	

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

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 72.563 on 56 degrees of freedom
Residual deviance: 14.384 on 52 degrees of freedom
AIC: 24.384

Number of Fisher Scoring iterations: 9

Annexe 53 : Effets marginaux moyens - modelef_logit

	pib	chomage	indice_gini	corruption
(Intercept)	5.073514e-01	1.521827e-05	1.888717e-02	-2.310281e-02
	6.450653e-02			

Annexe 54 : Résultats de la régression - modelef_probit

```
call:
glm(formula = electricite ~ pib + chomage + indice_gini + corruption,
     family = binomial(link = "probit"), data = pauvrete2, x = TRUE)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	7.7144803	3.4309280	2.249	0.02454	*
pib	0.0002332	0.0001086	2.148	0.03170	*
chomage	0.2841247	0.1511795	1.879	0.06019	.
indice_gini	-0.3509146	0.1335132	-2.628	0.00858	**
corruption	0.9691826	0.6469936	1.498	0.13414	

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

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 72.563 on 56 degrees of freedom
Residual deviance: 14.128 on 52 degrees of freedom
AIC: 24.128

Number of Fisher Scoring iterations: 11

Annexe 55 : Effets marginaux moyens - modelef_probit

	pib	chomage	indice_gini	corruption
(Intercept)	0.5008146126	0.0000151396	0.0184450287	-0.0227809446
				0.0629181510

Annexe 56 : Critères de comparaison logit / probit

```
> AIC(modelef_logit, modelef_probit)
      df      AIC
modelef_logit  5 24.38422
modelef_probit  5 24.12753
> BIC(modelef_logit, modelef_probit)
      df      BIC
modelef_logit  5 34.59948
modelef_probit  5 34.34279
> logLik(modelef_logit)
'log Lik.' -7.192111 (df=5)
> logLik(modelef_probit)
'log Lik.' -7.063765 (df=5)
> R2_Mc_Fadden_logit <- 1 - (modelef_logit$deviance/modelef_logit$null.deviance)
>
> R2_Mc_Fadden_logit
[1] 0.8017681
>
> R2_Mc_Fadden_probit <- 1 - (modelef_probit$deviance/modelef_probit$null.deviance)
>
> R2_Mc_Fadden_probit
[1] 0.8053057
```

Table des matières

Introduction	3
1. Présentation des données	4
A. Origine et caractéristiques des données	4
B. Présentation des variables par l'analyse documentaire	5
C. Résumé des variables du modèle	10
2. Analyse descriptive	11
A. Analyse univariée	11
1. Variables qualitatives	11
2. Variables quantitatives	12
a. Indicateurs de tendance centrale et de dispersion	12
b. Visualisation et traitement de valeurs atypiques	13
c. Caractéristiques des distributions	14
B. Analyse bivariée	16
1. Quanti-quanti	16
a. Nuages de points	16
b. Corrélations	17
2. Quali-quali	18
a. Tableau de contingence	18
b. Test de Khi-deux	18
c. V de Cramér	19
3. Quanti-quali	19
a. Tests de comparaison des moyennes	19
b. Statistique χ^2	19
3. ACP	21
A. Analyse simple	21
B. Analyse des clusters	23
1. Dimensions 1 et 2	23
2. Dimensions 3 et 4	26
4. Estimation économétrique et interprétation des résultats	29
A. Régression du modèle de base	29
1. Qualité du modèle (step)	29
2. Tests statistiques	29
a. Normalité des résidus	30
b. Forme fonctionnelle	30
c. Multicolinéarité	31
d. Homoscédasticité des résidus	31
e. Distance de Cooks	31
3. Interprétation des résultats	32
a. Test de Fisher	33
b. R^2	33
c. Test de Student	33
d. Interprétation des coefficients	34

B. Régression à partir de l'ACP	35
1. Choix et construction du modèle	35
2. Tests statistiques	35
a. Normalité des résidus	35
b. Forme fonctionnelle	36
c. Multicolinéarité	36
d. Homoscédasticité des résidus	36
3. Interprétation des résultats	37
a. Test de Fisher	37
b. R ²	37
c. Test de Student	37
d. Interprétation des coefficients	38
C. Comparaison des deux modèle	38
5. Modèles Logit / Probit	40
A. Modèle Logit	40
1. Recherche du meilleur modèle	40
a. Qualité du modèle (step)	40
b. Multicolinéarité	40
c. Test de Wald	41
2. Modèle logit final	41
3. Effets marginaux moyens	42
B. Modèle Probit	42
1. Modélisation	42
2. Effets marginaux moyens	43
C. Comparaison des modèles	44
Conclusion	46
Limites	48
Bibliographie	50
Annexe	52